Using Approximate Dynamic Programming to Solve the Military Inventory Routing Problem with Direct Delivery

THESIS

MARCH 2015

Rebekah S. McKenna, Second Lieutenant, USAF

AFIT-ENS-MS-15-M-140

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

The views expressed in this document are those of the author and do not reflect the official policy or position of the United States Air Force, the United States Department of Defense or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT-ENS-MS-15-M-140

USING APPROXIMATE DYNAMIC PROGRAMMING TO SOLVE THE MILITARY INVENTORY ROUTING PROBLEM WITH DIRECT DELIVERY

THESIS

Presented to the Faculty
Department of Operational Sciences
Graduate School of Engineering and Management
Air Force Institute of Technology
Air University
Air Education and Training Command
in Partial Fulfillment of the Requirements for the
Degree of Master of Science Opertaions Research

Rebekah S. McKenna, BS
Second Lieutenant, USAF

MARCH 2015

AFIT-ENS-MS-15-M-140

USING APPROXIMATE DYNAMIC PROGRAMMING TO SOLVE THE MILITARY INVENTORY ROUTING PROBLEM WITH DIRECT DELIVERY

Rebekah S. McKenna, BS
Second Lieutenant, USAF

Committee Membership:

Lt Col Matthew J. Robbins, PhD
Chair

LTC Brian J. Lunday, PhD
Member

AFIT-ENS-MS-15-M-140

Abstract

The United States Army uses Vendor Managed Inventory (VMI) replenishment to manage resupply operations while engaged in a combat environment; upper-echelon organizations (e.g., a brigade) maintain situational awareness regarding the inventory of lower-echelon organizations (e.g., battalions and companies). The Army is interested in using a fleet of cargo unmanned aerial vehicles (CUAVs) to perform resupply operations. We formulate an infinite horizon, discrete time stochastic Markov decision process model of the military inventory routing problem with direct delivery, the objective of which is to determine an optimal unmanned tactical airlift policy for the resupply of geographically dispersed brigade combat team elements operating in an austere, Afghanistan-like combat situation. An approximate policy iteration algorithm with Bellman error minimization using instrumental variables is applied to determine near-optimal policies. Within the least-squares temporal differences policy evaluation step, we use a modified version of the Bellman equation that is based on the post-decision state variable. Computational results are obtained for examples based on representative resupply situations experienced by the United States Army in Afghanistan.

Key words: inventory routing problem, Markov decision process, approximate dynamic programming

For my family, especially my grandmothers. Their unwavering support was invaluable.

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. J.D. Robbins, for his continuous patience, knowledge, and support during this effort. Without his guidance, this endeavor would not have been possible.

Rebekah S. McKenna

Table of Contents

Page

List of Figures

List of Tables

Table Page

USING APPROXIMATE DYNAMIC PROGRAMMING TO SOLVE THE
MILITARY INVENTORY ROUTING PROBLEM WITH DIRECT DELIVERY

I. Introduction

This thesis is motivated by the need to solve a military variant of the inventory routing problem (IRP). The inventory routing problem involves simultaneously making decisions about vehicle routing and inventory resupply, typically under vendor-managed inventory (VMI) practices [7]. Traditionally, a customer alerts a central vendor when resupply at its location is necessary. However, under VMI practices, a central vendor monitors the supply levels of its customers and chooses the amount of inventory to replenish at each location as well as the time at which to deliver it. Coelho *et al.* [7] describe inventory management practices as a "win-win," wherein the vendor benefits by realizing cost savings due to coordinating shipments and the customers benefit by no longer having to allocate resources to inventory management. When the inventory routing problem is solved, three key decisions are made at each decision epoch: which customers to serve, how much to deliver to each customer, and how to combine the deliveries to customers into optimal vehicle routes [7].

The IRP seeks decisions that minimize cost while satisfying a stipulated set of constraints. Many variations of the IRP exist. The elements which describe variations of the IRP include: time horizon, structure, routing policy, fleet composition, fleet size, supply assumptions, and demand assumptions [7]. The time horizon indicates the number of decision epochs considered in the problem and is chosen so as to reflect a planning horizon appropriate for the problem's context. The structure refers to the vendor/customer relationship and is typically modeled with one vendor and multiple customers. Routing can be either direct or indirect, indicating whether a vehicle can supply multiple customers before returning to the supplier. Fleet composition describes the type of vehicle(s) in the problem. A homogeneous fleet includes one type of vehicle, whereas a heterogeneous fleet has multiple types of vehicles. The fleet size refers to the number of available vehicles, either unconstrained or limited. The demand assumption refers to how demand is modeled in the problem. When stochastic demand is assumed, customers' demands are probabilistic in nature, whereas a deterministic demand indicates that the demand at each customer is known with certainty. The supply assumption parallels the demand assumption; it can be either stochastic or deterministic.

The United States Army utilizes vendor-managed inventory practices when engaged in combat operations to manage resupply operations. Upper-echelon organizations monitor inventory levels at lower-echelon organizations and manage resupply efforts for these lower-echelon organizations. Specifically, a brigade monitors and makes decisions regarding resupply efforts for battalions and companies. In Afghanistan, for example, a single brigade may manage resupply efforts for as many as 36 combat outposts (COPs), each of which serves as a base for a platoon- or company-sized infantry element.

COPs are traditionally resupplied by ground assets which are vulnerable to improvised electronic device (IED) attacks [18]. Development of increasingly capable cargo unmanned aerial vehicles (CUAVs) creates an alternative to ground resupply. While CUAVs present an opportunity to resupply COPs without endangering lives, the aircraft are vulnerable to hostile actions from non-friendly forces and can be lost during a resupply effort. The risk of losing a CUAV varies based on the time the CUAV is deployed and the location of the COP the CUAV is resupplying. The CUAVs are a limited commodity and cannot be used in future resupply operations once lost.

McCormack [18] introduces the military variant of the inventory routing problem (MILIRP) in which inventory is not guaranteed to reach its destination. Moreover, the MILIRP accounts for the possibilities of failed deliveries and lost vehicles due to hostile actions by non-friendly forces when a CUAV is enroute to its customer or returning to the depot. The long-term effect of the loss of a CUAV must be considered before choosing to send the CUAV on a resupply mission.

The MILIRP is formulated as a discrete time, infinite-horizon Markov decision process (MDP) with an objective of determining the optimal unmanned aerial vehicle resupply policy for the sustainment of brigade combat team elements operating in a combat environment. The optimal policy includes decisions regarding which COPs to resupply during specific time epochs. The MILIRP formulated in this thesis is classified as a one-to-many, direct routing problem with a homogeneous limited fleet composition. Considering the routing assumption particularly, direct delivery is assumed in this thesis. Direct delivery requires that every CUAV travels directly to a single COP and returns to the depot before visiting another COP. Therefore, optimization of vehicle routing will not be needed in the solution.

The MILIRP is formulated as a Markov decision process (MDP). When solving an MDP, an exact solution can frequently be found using backwards induction techniques. However, some MDPs are quite large and backwards induction techniques require long computation times to provide a solution. Instead, an approximate solution is found. An approximate solution, unlike an exact solution, may not be the optimal solution. Instead, the solution is the best policy found within a reasonable amount of time. For the MILIRP, the number of COPs is the variable which determines whether an approximate solution is needed. When the number of COPs increases above three, an approximate solution is necessary. For one, two, and three

COPs, an exact solution can be determined using backwards induction. Therefore, a *small instance* of the MILIRP refers to instances having three or fewer COPs, whereas a *large instance* of the MILIRP refers to instances having greater than three COPs.

Using an exact dynamic programming algorithm (e.g. policy iteration), small instances of the problem with up to three COPs are solved. The necessity of an approximate dynamic program (ADP) is demonstrated using a larger number of COPs. Then, by comparing the approximate solution to the exact solution for small instances of the problem, the quality of the ADP can be assessed. To maximize the performance of the ADP, the parameters of the ADP are modified. Once these parameters are developed, a larger instance of the MILIRP with up to 18 customers is approximately solved using the ADP. Sensitivity analysis is then performed.

The remainder of this thesis is organized as follows. Chapter 2 provides a review of literature pertinent to the MILIRP. This review includes discussion on the inventory routing problem and a review of the background for the MILIRP as provided by McCormack [18]. An introduction to Markov decision processes as well as approximate dynamic programming is provided. The methodology for finding an approximate solution to the MILIRP is considered in Chapter 3. Moreover, a formulation of the MDP is presented, as well as the ADP algorithm. The fourth chapter summarizes the analysis and results of the research, to include an interpretation of the optimal policy as well as a comparison of the exact solution to the approximation. The 12-COP problem is discussed and sensitivity analysis is provided. Chapter 5 presents conclusions regarding the results of the thesis and suggests further applications and research on the MILIRP.

II. Literature Review

The inventory routing problem (IRP) is a thoroughly researched topic in the operations research field because many industries rely on the transportation and management of goods. To aid in understanding the formulation and techniques for solving the military inventory routing problem (MILIRP), the literature review describes three main topics: the inventory routing problem, the military inventory routing problem, and Markov decision processes (MDP). The original contribution of this thesis is then discussed.

2.1 Inventory Routing Problem

The MILIRP is a variant of the inventory routing problem. Coelho *et al.* [7] provide a summary of the literature on the IRP. The objective of an IRP is to minimize the total cost to the supplier while meeting the demands of the customers, subject to the following constraints. First, the inventory cannot exceed the maximum capacity at each customer. Additionally, inventory levels cannot be negative. Third, vehicles must start and end their routes at the supplier, and each vehicle can deliver only once per time period. Finally, a vehicle's capacity cannot be exceeded [7].

The IRP seeks to provide answers to three questions: (1) in which time periods should each customer be served, (2) what amount of supplies should be delivered to each of these customers, and (3) how should customers be combined into vehicle routes [7]. Coelho *et al.* [7] and Kleywegt *et al.* [12] identify structural components which can be used to describe the variants of IRPs. These characteristics include: time horizon, structure, routing, inventory policy, fleet composition, fleet size, and demand type. The MILIRP is formulated as an infinite time horizon, one-to-many structured, direct routing, homogeneous fleet composition, limited fleet size, deterministic demand, and stochastic supply problem. To inform the formulation of the MILIRP, IRPs with similar structures are examined. Table 1 summarizes similar IRP formulations found in the sources (where D indicates *deterministic* and S indicates *stochastic*).

Table 1. IRP Formulations

Reference	Routing	Fleet size	Demand	Supply	Solution
Bertazzi (2008)	Direct	Unconstrained	D	D	Link Optimization
Barnes-Schuster & Bassok (1994)	Direct	Unconstrained	D	S	Simulation
Kleywegt et al. (2002)	Direct	Limited	S	D	ADP
Kleywegt et al. (2004)	Multiple	Limited	S	D	ADP
MILIRP	Direct	Limited	D	S	ADP

Direct delivery simplifies the IRP by removing the routing portion of the problem. With direct delivery, a vehicle moves from the supplier to the customer and returns immediately to the supplier without stopping at other customer locations. Bertazzi [3] shows that routing can significantly reduce the cost for particular classes of prob–lems. He also identifies other classes of problems in which direct delivery should be used, such as when the capacity of the vehicle roughly equals the demand of a customer [3]. In a later work, Bertazzi *et al.* [4] apply a rollout approach to solve the IRP while considering stock-out. Barnes-Schuster & Bassok [1] develop a similar one-to-many problem with direct delivery, but they assume stochastic demand. The authors establish that, when a normal distribution can be used to estimate demand, a direct routing policy should be employed if vehicle capacity is near the mean of customer demand [1].

Coelho *et al.* [7] provide an introduction to the stochastic inventory routing prob–lem (SIRP) wherein customer demand is stochastic. Shortages of supply are usually discouraged using penalty functions based on the amount of unfilled demand. Some problem formulations allow backlogging, i.e. allowing a supplier to fulfill unsatisfied demand in later periods. Given the stochastic nature of these problems, the goal of the IRP is to determine a policy which maximizes the expected discounted value over the time horizon of the problem [7]. Coelho *et al.* [7] identify three methods for solving the SIRP: heuristic algorithms, dynamic programming, and robust optimization.

Kleywegt *et al.* [12] form a direct-delivery stochastic inventory routing problem as an MDP. In particular, the states of the system are the inventory levels at each customer, and the action space includes the amount of inventory delivered to each customer. The state at epoch $t+1$ is dependent on the amount of inventory delivered, probabilistic demand, and the supply capacity of the customer during epoch t. Contributions are based on the traveling costs of the vehicles, shortage costs, holding costs, and revenue. Given the large state space, an approximate dynamic programming (ADP) algorithm is developed. Kleywegt *et al.* [12] provide an approach to solving the IRP with direct delivery and stochastic demand for an infinite horizon problem with homogeneous vehicles and no backlogging.

Kleywegt *et al.* [13] extend Kleywegt *et al.* [12] by removing the direct delivery constraint; a vehicle can make up to three stops at different customers before returning to the supplier. Relaxation of the direct delivery constraint requires the consideration of larger state and action spaces to account for available routes and assignment of routes to each vehicle. Due to the large size of their problem, an exact solution to the MDP is computationally infeasible. Further development of the ADP from Kleywegt *et al.* [12] is used to determine an approximate policy.

Coelho & Laporte [6] develop exact solutions for several classes of the IRP including the multi-vehicle IRP with homogeneous and heterogeneous fleets. The authors add additional features to the problem to address workforce management and regularity of service concerns [6]. These features include quantity consistency, vehicle filling rate, order-up-to level, driver consistency, driver partial consistency, and visit spacing. These features are implemented as constraints or as penalties in the ob–

jective function. Coelho & Laporte [6] use a a branch-and-cut algorithm to find the exact solutions to the problem classes.

The vehicle routing problem (VRP) is a component of the inventory routing problem; literature on this subject offers insight into this element of the problem. The vehicle routing problem with stochastic demand (VRPSD) gives particular attention to problems with stochastic demand. The VRPSD assumes that customers' demands are stochastic in nature, and that the true demand is realized only after a customer is supplied. Novoa & Storer [20] formulate a single-vehicle problem in which the objective is to determine a routing policy to minimize transportation costs and satisfy demand at each customer. An initial route is followed, but if demand cannot be filled during a single trip, a vehicle must return to the supplier to retrieve additional supplies. The authors use a Monte Carlo simulation to reduce computation time for a rollout algorithm. Through the improvement of the rollout-algorithm, Novoa & Storer [20] provide an efficient dynamic approach to solving the VRPSD for one vehicle.

A particular nuance of the MILIRP is that vehicles can be destroyed while traveling to and from the supplier which imposes a stochastic nature on the supply. Vehicle routing problems with vehicle breakdown have a similar complexity in a civilian context. Mu *et al.* [19] solve a variant of the VRP in which a new routing solution must be created in the event of a vehicle breakdown. The authors develop two metaheuristic which focus on rescheduling the route in an allotted time with a single extra vehicle available for use in the event of a breakdown. However, the Mu *et al.* [19] formulation differs fundamentally from the MILIRP in that the authors solve the re-optimization in only a single time period. The MILIRP must be solved over an infinite time horizon.

2.2 The Military Inventory Routing Problem

The IRP provides a starting basis for formulating and solving the MILIRP. Mc–Cormack [18] provides an introduction to the Army's resupply policies and practices. Sustainment operations allow the Army to extend its operational range for longer periods of time, a key to the Army's success [18]. McCormack [18] identifies four principles of sustainment pertinent to the MILIRP: 1) responsiveness, 2) simplicity, 3) economy, and 4) survivability. The combination of these principles allows com–manders to create timely, consistent, financially feasible, and smart policies regarding resupply efforts.

The MILIRP is motivated by the need to sustain subordinate and geographically disparate elements within an infantry brigade combat team (IBCT) in an austere combat environment. An IBCT is responsible for the supply of combat outposts (COPs) in its area of operation (AO). The relationship between an IBCT and a COP parallels the supplier-to-customer relationship seen in vendor managed inventory replenishment practices.

An IBCT contains a brigade support battalion (BSB) responsible for the resupply of COPs within its AO. A supply officer within the BSB is responsible for planning all sustainment efforts. This job entails coordinating and monitoring subordinate units' supply needs. A General Dynamics report found that an infantry company requires 25,000 lbs of supplies per day, and it divides the supplies into six categories: subsistence, construction items, ammunition, medical supplies, repair parts, and fuel [28]. The BSB is kept informed of inventory levels at COPs through regular reporting. This vendor managed inventory practice allows the IBCT to choose when and where to send supplies [18].

Resupply efforts pose a significant risk to personnel in combat environments. The Army typically operates in harsh, rugged environments that often include moun–

tains, deserts, and jungles [18]. Army resupply efforts are traditionally heavily reliant on ground lines of communication (GLOC) [18]. However, a lack of transportation infrastructure and/or attacks from the enemy make GLOC resupply inherently dangerous and difficult. Improvised explosive devices (IED) accounted for “65% of U.S. deployed fatalities between November 2002 and March 2009, with 18% occurring during sustainment operations” [10].

Despite these challenges, resupply efforts continue. Manned air assets partially fill the resupply role but also have limitations. Pilots cannot fly in hazardous weather conditions, and helicopters are vulnerable to man-portable air defense systems (MANPADS), especially during takeoff and landing at the COPs. While some manned helicopters have armed escorts to mitigate the MANPADS risk, resupply missions are canceled when the threat of attack is too high [18]. Additionally, with a limited supply of air assets and a high operational tempo, McCormack [18] indicates that supporting combat missions is prioritized over resupply efforts. Contractors are hired for aerial resupply efforts, but strict constraints (including eight hour shift limits and a ban on nighttime sorties) limit their ability to alleviate the strain on military helicopters [18]. McCormack [18] concludes that ground resupply convoys are still necessary, given the limitations to manned aerial resupply and constraints on contractors. An infantry officer underscored the lack of flexibility with the current system when he said, “If the support is not anticipated (more than 72 hours out), then you are not getting the support. There is no immediate resupply” [10].

Due to resupply challenges, the Army is considering the use of rotary cargo unmanned aircraft vehicles (CUAVs) for use in sustainment efforts. Williams [28] explores the use of unmanned airlift at both the theater and direct delivery levels in Department of Defense applications. Motivation to use CUAVs is prevalent, from commanders in the field to Congressional representatives. Williams [28] indicates that field units frequently request unmanned aerial systems and specifically identifies forward operating base support as an area were CUAVs could make a dramatic im–pact. Congress is also interested in increasing the role of the unmanned aerial vehicles (UAVs). The DOD's *Unmanned System Integrated Roadmap* provides an outline of strategic and tactical airlift goals through 2028, and it establishes the potential role of unmanned aerial and ground systems [27]. In particular, the Army sponsored a General Dynamics study on unmanned aircraft in resupply roles [10]. The study recommended that the CUAV system be centrally managed, rather than dedicated to a BSB, to allow for increased effectiveness in using the CUAV for multiple roles. Without an exclusive resupply mission, the CUAVs may be plagued by similar issues as manned rotary aircraft in balancing support and combat missions.

McCormack [18] identifies a number of benefits to utilizing CUAVs. First, a dedicated contingent of CUAVs for resupply would free manned helicopters for com–bat mission use and mitigate the risks associated with GLOC resupply. The CUAV's higher flight ceiling and better performance in adverse conditions would reduce MAN–PADS threats, possibly allow for shorter supply routes, and provide a quicker, more reliable, and more flexible delivery platform. However, some challenges must be ad–dressed if CUAVs are used in a resupply role: 1) demand for large quantities of supplies across area of operations, 2) effects of enemy threat and action, 3) weather, terrain, and poor infrastructure, 4) availability of distribution assets, and 5) flexibility to respond to changes in the operational environment [18].

Williams [28] identifies two CUAVs in development that could meet the demands of a tactical airlift role: the Boeing A160 Hummingbird and the Lockheed Martin Kaman K-MAX. Williams [28] reports that the two CUAVs in development (the Hummingbird and the K-MAX) met DoD requirements during testing. The author also indicates that the aircraft have the capability to be used in the combat zone.

In fact, three K-MAX vehicles were deployed to Afghanistan between 2011 and 2014 [17]. As motivation at both the tactical and strategic level for developing CUAVs continues to grow, technical development of these air systems makes steady progress.

Barriers to CUAV development remain. Williams [28] identifies two key safety issues. First, the CUAV must be able to recognize objects in the aircraft's flight path and reroute around the obstacle. Moreover, command and control of the aircraft must be guaranteed. Without these safety measures, CUAVs may crash or fall into the hands of the enemy.

With political and military support of the CUAV growing and technological development nearly complete, the CUAV has the potential to positively impact the Army's resupply efforts. However, decision makers would benefit from a method for managing this new platform. Without a mechanism for determining how to allocate the CUAVs for resupply, the CUAV could be non-optimally utilized. The goal for the MILIRP is to suggest a policy for utilizing this resource. McCormack [18] formulates and solves the MILIRP for a single COP exactly. The original contribution of this thesis is approximately solving the MILIRP for a large number of COPs using an approximate dynamic program.

2.3 Markov Decision Processes

Formulating the MILIRP as a Markov decision process (MDP) provides the foundation for structuring the problem. With this structure in place, solutions to the MILIRP can be found both exactly and approximately. Puterman [22] introduces MDPs by describing a decision maker who must make decisions at discrete points as a system evolves over time. The decision maker chooses an action based on the current state of the system. Once an action is chosen, the system evolves either deterministically or stochastically, and an immediate expected reward is gained. The system then arrives in a new state, and the decision maker repeats the process of choosing an action, receiving a reward, after which the system transitions to a new state. When considering systems with a finite set of decision epochs, a final reward may be also be realized. The goal of an MDP is to create a policy of decision choices which maximizes the reward a decision maker receives over the lifetime of the system. Key to this process is the idea that future consequences must be accounted for in earlier decisions [22].

Puterman [22] outlines the five elements of an MDP which form the structure of the problem: states, actions, the time horizon, transition probabilities, and rewards. The *state* of the system is a description of the elements of the system used to make future decisions [22]. A state space is a set of the possible states the system can occupy. An *action* describes alternatives a decision maker can choose in a particular state. An action space is a set of the possible actions a decision maker can choose at a particular state. A MDP can be modeled with either an infinite or a finite *time horizon*. The *transition probabilities* describe the probability a system will transfer from one state to another state given a particular action. Finally, the *reward* is either an immediate reward or a terminating reward. This reward is also referred to as a *contribution*. The actions taken based on the rewards and transition functions can only depend on the state of the system during the current time period. Moreover, the previous states of the system cannot affect the current decision, a key assumption for using an MDP [22].

The following notation is used to formulate an MDP. Let $\mathcal{T}$ denote the discrete set of points in time (i.e. epochs) at which decisions are made. At time $t \in \mathcal{T}$, the system is in state $s_t \in \mathcal{S}$, where $\mathcal{S}$ is the set of all the possible states of a system. The decision maker chooses action $a \in \mathcal{A}_s$, where $\mathcal{A}_s$ is the set of all possible actions. The decision maker then receives an immediate contribution, $r(s, a)$. The transition function describes the probability that the system transfers to state j from state s given action a is taken: $p(j|s,a)$. An optimal decision rule, $d(s)$, describes the action the decision maker makes in state s. Once formulated, an MDP can be solved using Bellman's optimality equation, shown in Equation 1, where $J(S_t)$ denotes the value of being in state S_t at time t and λ represents the discount factor. The discount factor indicates the present value of a single unit of the reward if it were recieved at the next time period [22].

$$J(S_t) = \max_{a \in \mathcal{A}(S_t)} (r(S_t, a) + \lambda \mathbb{E}\{J_{t+1}(S_{t+1})|S_t\}) \tag{1}$$

Bellman's equation provides a mechanism for obtaining the exact solution to an MDP, where a decision rule for each time epoch is determined. However, in many cases the problem may not be computationally solvable due to the *curse of dimensionality.* This limitation occurs when a problem's state space or action space becomes too large or the transition function becomes too complex to specify. In these instances, it may take years for a computer to arrive at a solution. To avoid this problem, the field of approximate dynamic programing (ADP) seeks to approximate the value function. This approach provides an approximate solution to the problem rather than an exact solution [23]. Powell [21] provides a standard for formulating and solving approximate dynamic programs. To find an exact solution, backward induction is used to recur–sively compute the expectation of the discounted value function over all states and actions. Instead, ADP relies on forward induction using simulation to approximate the value function of the problem. The quality of the approximate solution can be estimated by solving a small instance of a larger problem and comparing the exact solution found using Bellman's equation with the approximate solution found using a value function approximation.

Approximate value iteration (AVI) provides a method for solving ADPs by simul–

taneously updating the value function approximation and policy approximation [21]. This approach approximates the value function by sampling states randomly and re–peatedly approximating the value function over numerous iterations. The function eventually converges to the correct value without looping through every state and action. This approach is explored in the reinforcement learning community with *Q-learning*, where the value of the state-action pair is estimated (rather than just the state). Tsitsiklis & Sutton [26] provide a proof of convergence for Q-learning using lookup table representations. However, this method cannot be used for large scale problems given the need for a lookup table.

Approximating the value function using AVI is explored in depth by Topaloglu & Powell [25], who exploit the concave value function structure of stochastic resource allocation problems. Topaloglu & Powell [25] present the leveling algorithm which builds piecewise linear approximations of the value function by sampling the gradient of the function and maintaining concavity at each iteration. Using this algorithm, slopes which violate monotonicity are updated by sampling a stochastic outcome and determining the sample gradient information. If monotonicity in the slope is not maintained, the slopes are leveled to equal the current slope's value. After a number of iterations, the value function is approximated. This value function approximation is then smoothed; the old value estimate is combined with the new estimate using an α value, also called a stepsize. The stepsize determines the rate at which the new estimate is combined with the old estimate [21]. Topaloglu & Powell [25] provide proof of convergence for the leveling algorithm indicating that, after a sufficient number of iterations, the approximation will approach the exact solution. Godfrey & Powell [11] use the same leveling algorithm to maintain monotonicity but include a dynamic allocation of breakpoints to create the CAVE algorithm. AVI is limited in its ability to provide a general solution for all problem classes.

Approximate policy iteration (API) provides a solution algorithm that does not depend on approximating a state-action pair (as in Q-learning) and possesses a strong convergence theory [23]. Bellman's equation is modified to represent an approximation of the value function based on the *post-decision state*. The post-decision state variable, S_t^a, is the state the system is in once an action has been taken, but prior to any exogenous information being realized [23]. The value of the post-decision state, $J^a(S_t^a)$, is the value of the system existing in the post-decision state. The value function for the post-decision state is therefore the expected value of being in a state at the next time epoch given the system is currently in a post-decision state, as shown in Equation 2.

$$J_t^a(S_t^a) = \mathbb{E}\Big\{ \max_{a \in \mathcal{A}(S_{t+1})} (r(S_{t+1}, a) + J_{t+1}^a(S_{t+1}^a)) | S_t^a \Big\} \tag{2}$$

Bellman's equation around the post-decision state variable must be approximated. Bradtke *et al.* [5] introduce the least squares temporal difference (LSTD) method which estimates the value of a fixed policy. Within a machine learning context, temporal difference algorithms allow a system to learn to predict the results of decisions over a specific time horizon. The modification of temporal difference algorithms to include least squares provides a more efficient use of sample data [14]. Bradtke *et al.* [5] provide two temporal difference algorithms (i.e. a least-squares and recursive least squares) and conclude that the rate of convergence for LSTD is faster than basic temporal difference techniques.

Lagoudakis & Parr [14] introduce least squares policy iteration (LSPI). First the authors define LSTDQ, an algorithm which approximates the state-action value function, bypassing the need for a transition function. Lagoudakis & Parr [14] then apply LSTDQ within a policy iteration algorithm framework to create LSPI. These two techniques combine efficient policy-search algorithms with efficient use of sample data

[23]. LSPI is able to overcome a central issue with LSTD by not being affected by the number of times a state is visited [23]. Lagoudakis & Parr [14] build on this method by incorporating the use of instrumental variables (IV) into the LSTD method to mitigate errors in approximating the independent variables due to correlations with the error term. The method is coined the least squares approximate policy iteration (LSAPI).

Scott *et al.* [23] contribute to LSAPI research by utilizing least-squares Bellman error minimization. Powell [21] explains Bellman error as the temporal difference which reflects the difference in estimating the value of being in a state at time t at the current iteration and the updated iteration. Using Bellman error minimization, Scott *et al.* [23] provide three augmentations to LSAPI, one of which utilizes instrumental variables. The authors use this algorithm to solve an energy storage problem and show that LSAPI using instrumental variables outperforms basic LSAPI.

III. Methodology

3.1 Problem Description

McCormack [18] proposes the military inventory routing problem (MILIRP). A notional infantry brigade combat team (IBCT) is responsible for a number of combat outposts (COPs) within its area of operations (AO). The IBCT contains a brigade support battalion (BSB) which manages resupply efforts for G number of COPs. The BSB manages V number of identical CUAVs which deliver supplies to the COPs. Each CUAV has a load capacity of Q pounds and it is assumed that each CUAV is fully loaded when dispatched. COP i requires d_i pounds of supplies per time period, a deterministic demand which depends on the size of the unit at the COP. Only direct deliveries are considered; each CUAV visits only one COP per trip. This formulation reduces the complexity of the problem and reflects the fact that current rotary assets cannot combine multiple deliveries [18].

Given the austere combat environment, there is a potential for delivery failure due to factors such as hostile actions by non-friendly forces, mechanical failures, and extreme weather conditions. McCormack [18] proposes a tessellation of the AO in which each hexagonal cell is identified as a high or low threat area. The probability of a CUAV being destroyed depends on the COP being resupplied and the current threat map. A set of K threat maps is created to reflect the periodic changes in risk for an AO. Dijkstra's algorithm is applied to determine an optimal path from the BSB to each COP i for each tessellated threat map K. Associated with each optimal path is ψ_{ik}, the probability of successfully completing a one-way trip from the BSB to COP i (and from COP i to the BSB) under threat conditions indicated by map K.

The CUAV has two opportunities to be destroyed: either traveling to the COP from the BSB or returning to the BSB after delivering supplies to the COP.

With this background, the MILIRP is formulated as a Markov decision process (MDP). When considering imposing an inventory routing formulation on the MILIRP, the CUAVs are the vehicles, the COPs are the customers, and the BSB is the supplier. Table 2 provides a summary of notation at the end of this chapter.

3.2 MDP Formulation

The MDP formulation includes the following components: a time horizon, states, actions, transition probabilities, rewards, and an objective function. A finite number of CUAVs are available at the BSB. The BSB knows the inventory level of each COP i at any time t. At each decision epoch, the number of CUAVs deployed to each COP is determined. An immediate reward is gained if a CUAV successfully reaches the COP. If a CUAV fails to return to the BSB, that CUAV is considered non-operational and cannot be used in future CUAV resupply missions. If a COP depletes its supplies, a penalty is applied. Once all CUAVs are non-operational, the system has evolved into an absorbing state. The objective is to determine a deterministic, stationary policy which maximizes the expected total discounted reward.

Time Horizon

The MILIRP is formulated with an infinite time horizon, $t \in \mathcal{T} = \{1, 2, ...\}$. During a single time period a CUAV is fueled, loaded with supplies, travels from the BSB to the COP, unloads its supplies, and returns to the BSB. It is assumed that a fully loaded CUAV can serve each COP within the AO during this time period. At each decision epoch, the number of CUAVs deployed to each COP is determined.

States

The state $S = (x_1, x_2, ..., x_G, v, k) \in \mathcal{S}'$ captures the current amount of inventory at G COPs, the current number of operational CUAVs, and the current threat map.

Let C_i denote the the inventory capacity at COP i. Therefore, the state space is $\mathcal{S}' = [0, C_1] \times [0, C_2] \times ... \times [0, C_G] \times \{1, 2, ..., V\} \times \{1, 2, ..., K\}$ if the amount of supplies is continuous or $\mathcal{S}' = \{0, 1, ..., C_1\} \times \{0, 1, ..., C_2\} \times ... \times \{0, 1, ..., C_G\} \times \{1, 2, ..., V\} \times \{1, 2, ..., K\}$ if the amount of supplies is discrete. Let $x_{it} \in [0, C_i]$ (or $x_{it} \in \{0, 1, ..., C_G\}$) denote the inventory level of COP i at time t. Let $v_t \in \{1, 2, ..., V\}$ denote the number of operational CUAVs available at time t. Let k_t denote the map at time t. The dimensionality of the state space is $G+2$, depending on the number of COPs investigated in the problem. The full state space $\mathcal{S} = \mathcal{S}' \cup \{\triangle\}$ consists of $\mathcal{S}'$, augmented by $\triangle$, the absorbing state. The state $S = \triangle$ denotes the situation where no CUAVs are operational; this occurs when $v_t = 0$. Let $S_t \in \mathcal{S}$ denote the state of the system at time t.

Actions

Let $\mathcal{A}(S)$ denote the set of all feasible decisions when the system is in state S. A decision $a = (a_1, a_2, ..., a_G) \in \mathcal{A}(S)$ denotes the number of CUAVs deployed to each COP. During a single epoch, a CUAV travels directly from the BSB to a COP, unloads, and returns directly back to the BSB. Two constraints are placed on the number of CUAVs which can be deployed at each time epoch. First, the number of CUAVs deployed cannot exceed the number of operational CUAVs, v_t. Second, the total number of CUAVs deployed cannot exceed the number of operator crews available in the BSB, κ. Let $a_t = (a_{1t}, a_{2t}, ..., a_{Gt}) \in \mathcal{A}(S)$ denote the decision made at time t when the system is in state S, where $a_{it} \in \{0, 1, ..., \min(v_t, \kappa)\}$ is the number of CUAVs sent to resupply COP i at time t.

$$a_t = A^{\pi}(S_t|\theta) \in A(S_t) \tag{3}$$

Moreover, the following constraint must hold:

$$\sum_{i=1}^{G} a_{it} \leq \min(v_t, \kappa), \forall\, t \in \mathcal{T}. \tag{4}$$

Transition Probabilities

Transition probabilities are defined for each dimension of the state space including the inventory levels at each COP, the number of CUAVs, and the threat map.

The inventory transitions are based on the amount of supplies gained by each COP at time t, which depends on the routing decision, a_t, and the state of the system, S_t. There are three possible outcomes when a CUAV is deployed on a resupply mission to a COP: 1) successful delivery to the COP and successful return to the BSB (denoted as an SS event), 2) successful delivery to the COP but failure to return to the BSB (denoted as an SF event), and 3) failure to arrive at the COP enroute from the BSB (denoted as an F event). Let $\psi_{ik}^2, \psi_{ik}(1-\psi_{ik})$, and $(1-\psi_{ik})$ denote the probabilities associated with an SS, SF, and F events occurring, respectively, when a single CUAV is sent to resupply COP i during threat conditions indicated by map k. The outcome of a resupply decision involving multiple CUAVs delivering supplies to a particular COP can be represented using the multinomial distribution. However, since we are interested in the specific class of outcome (i.e., SS, SF, or S), we proceed by defining the marginal distributions for each class, which are all binomial. We assume the outcome of each resupply mission to a particular COP is independent of the outcome of the resupply missions to other COPs. Let $Z_{it,SS}|S_t, a_{it}$ denote the number of possible successful deliveries made to COP i with a successful return to the BSB (i.e., an SS event), during time epoch t, on map k_t with a_{it} CUAVs deployed. The random variable $Z_{it,SS}$ follows a binomial distribution with parameters a_{it} and ψ_{ik}^2. Let $Z_{it,SF}|S_t, a_{it}$ denote the number of possible successful deliveries made to COP i with an unsuccessful return to the BSB (i.e., an SF event), during time period t, on map k_t with a_{it} CUAVs deployed. The random variable $Z_{it,SF}$ follows a binomial distribution with parameters a_{it} and $\psi_{ik}(1-\psi_{ik})$. Let $Z_{it,F}|S_t, a_{it}$ denote the number of possible failed deliveries made to COP i (i.e., an F event), during time period t, on map k_t with a_{it} CUAVs deployed. The random variable $Z_{it,F}$ follows a binomial distribution with parameters a_{it} and $1-\psi_{ik}$.

The amount of supplies delivered to COP i at time t is $Q(Z_{it,SS}+Z_{it,SF})$, where Q is the CUAV's capacity. There is a constraint placed on the amount of supplies which can be delivered to each COP; a CUAV's delivery cannot result in the COP exceeding its capacity, C_i. Moreover, if a COP's inventory reaches zero, the COP is immediately resupplied to capacity by ground lines of communication (GLOC). Since all COPs are accessible via ground infrastructure, the assumption that GLOC resupply is available at all COPs is realistic. However, the assumption that a commander would wait until a COP is completely depleted to order a ground resupply is not realistic. Additionally, assuming that all GLOC resupplies are success is not valid due to the poor transportation infrastructure and the enemy actions previously discussed. Despite these concerns, the GLOC assumptions used in this thesis are necessary to adequately model the problem: GLOC missions ensure COPs are resupplied when CUAVs are not available. Equation 5 is the inventory transition function for COP i.

$$X_{i,t+1} = \begin{cases} C_i & \text{if } X_{it} + Q(Z_{it,SS}+Z_{it,SF}) - d_i < 0, \\ \min(X_{it} + Q(Z_{it,SS}+Z_{it,SF}) - d_i, C_i) & \text{otherwise.} \end{cases} \tag{5}$$

In the first case, GLOC resupply is necessary and the COP is resupplied to capacity. In the second case GLOC resupply is not necessary and a minimization enforces the capacity constraint.

The transition function for the number of operational CUAVs is based on the probability a CUAV successfully travels between the BSB and the COP. If a CUAV fails enroute to the COP or returning to the IBCT (i.e., events SF or F), the CUAV is lost and cannot be used in future resupply efforts. The vehicle transition function is given in Equation 6.

$$v_{t+1} = v_t - (Z_{it,SF} + Z_{it,F}) \tag{6}$$

The map transition function represents the evolution of an uncontrolled, stochastic aspect of the operational environment. The set of maps captures the threat environ–ment. Whereas some maps represent a low threat environment with a low number of tessellated regions labeled as high threat, other maps present a high threat envi–ronment with a high number of tessellated regions labeled as high threat. As more tessellated regions are labeled as high threat, delivery of supplies using the CUAVs becomes increasingly risky. Different CUAV routes are used for each particular threat map; recall that we apply Dijkstra's algorithm to each COP i for each map k *a priori* to solving the MDP, which allows us to find the route with the highest one-way prob–ability of survival, ψ_{ik}. An increasing number of high threat tessellated regions may result in much lower one-way probabilities of survival, depending on the location of the BSB, COP, and the high threat regions.

The map transition represents the probability of the threat environment changing. If the operational environment is relatively static, the transition probabilities between maps would be relatively low. If the operational environment changes rapidly between high and low threats, the transition probabilities would be relatively high. It is con–ceivable that the transition probabilities could be constructed in a variety of ways. For example, IBCT intelligence teams working with the BSB may be able to use risk assessments to label a tessellated region based on information such as enemy dis–position, weather, and season. Specifically, low winds are particularly important to successful rotary aircraft flight. Information about mechanical failures of the CUAVs or operational crew reliability may also be captured in this risk assessment. Alterna–tively, historical data from enemy engagements and weather conditions could be used to label the threat map in a similar manner.

The outcome of the resupply missions, Z_t, provides the source of randomness in the MDP where $Z_t = (Z_{1t}, Z_{2t}, ..., Z_{Gt})$. The known joint probability distribu–tion, H, of inventory, vehicle, and map transitions gives a known Markov transition function W, according to which transitions occur. For any state $S \in \mathcal{S}$, any ac–tion $A \in \mathcal{A}(S)$, and any Borel subset $B \subseteq \mathcal{S}$, let $\mathcal{Z}(S, A, B) \equiv \{Z \in \mathbb{R}_+^G \times \mathbb{Z}_+^Z : (X_{1,t+1}, ..., X_{G,t+1}, v_{t+1}, k_{t+1}) \in B\}$. Then $W[B|S, A] \equiv H[\mathcal{Z}(S, A, B)]$. In other words, for any state $S \in \mathcal{S}$, and any action $A \in \mathcal{A}(S)$, $P[\mathcal{S}_{t+1} \in B|\mathcal{S}_t = S, \mathcal{A}_t = A] = W[B|S, A] \equiv H[\mathcal{Z}(S, A, B)]$.

Contribution

The contribution function is defined by the amount of supplies delivered to each COP. If the amount of supplies delivered, $Q(Z_{SS} + Z_{SF})$, results in an inventory at COP i which exceeds its capacity, C_i, only the amount of supplies up to the capacity is included in the reward. An immediate cost is applied when stocking out occurs. Let τ_i represent the cost of stock out at COP i. Different penalties can be used to capture the difficulty of resupplying particular COPs via GLOC. COPs with higher penalties would receive more attention. The contribution function is presented in Equation 7.

$$r(S_t, a_t) \equiv \sum_{i=1}^{G} \min\left(C_i - X_{it} + d_i, Q(Z_{it,SS}|a_{it} + Z_{it,SF}|a_{it})\right) - \sum_{i=1}^{N} \tau_i I_{\{X_{i,t+1}<0\}} \quad (7)$$

The amount of deliverable supplies is determined by taking the minimum of the available capacity at COP i and the number of supplies delivered to COP i. No reward is gained for any supplies which would force the COP to exceed capacity. The indicator variable, I, equals one if the system is in a state where inventory is depleted and zero otherwise. This allows for a penalty to be applied when GLOC resupply is necessary.

Value Function

The objective of the MDP is to maximize the expected total discounted value over an infinite horizon. As with all Markov decision processes, the decisions made at time t depend only on the current state of the system, and the decision maker does not know what will happen in the future. To obtain a policy that maximizes the expected total discounted reward over all $t \in \mathcal{T}$, Bellman's equation is used:

$$J(S_t) = \max_{a \in \mathcal{A}(S_t)} (r(S_t, a) + \lambda \mathbb{E}\{J_{t+1}(S_{t+1})|S_t\}) \tag{8}$$

Using this MDP formulation, a dynamic programming algorithm is developed to obtain an optimal policy for CUAV resupply.

3.3 Exact Solution

Two methods for solving MDPs, value iteration and policy iteration, are popular due to their ease of implementation and strong convergence rates. The value iteration algorithm estimates the value of a pre-decision state given a current expectation of contributions and the expected value of possible outcomes [21]. Using an error tolerance parameter, the algorithm continues until a convergence criterion is met. The resulting set of actions comprises an optimal policy within a finite number of iterations [21] [22]. Although different variants of this algorithm exist, we look instead to the second algorithm, policy iteration, to solve the MILIRP due to its ease of implementation on infinite horizon problems.

Using the policy iteration algorithm, small instances of the MILIRP are solved, including the 2- and 3-COP problems. The optimal policy for these problems is then used as a comparison to the approximate policies obtained for the 2- and 3-COP problems. This comparison allows the ADP parameters to be tuned to create an algorithm which provides the best solution for small instances. Furthermore, a myopic solution is also considered which is used to evaluate the fidelity of the ADP for large problem instances.

3.4 Approximate Solution

The approximate dynamic program (ADP) developed applies the least squares instrumental variables approximate policy iteration (IVAPI) algorithm. The pseudo code for the IVAPI algorithm is shown in Algorithm 1. The IVAPI methodology fundamentally relies on three ideas: approximate policy iteration (API), least squares temporal differences, and Bellman error minimization using instrumental variables.

Approximate Policy Iteration

Both approximate policy iteration and approximate value iteration mirror their exact counterparts as the two most widely used methodologies for discovering near-op–timal policies using value function approximation. Instead of using a one-step transi–tion matrix to solve the MDP exactly, the value function must be approximated and updated. Looking specifically to API, the value function is approximated around the post-decision state to avoid the need to find the exact value function.

$$J_t^a(S_t^a) = \mathbb{E}\Big\{ \max_{a \in \mathcal{A}(S_{t+1})} (r(S_{t+1}, a) + J_{t+1}^a(S_{t+1}^a)) | S_t^a \Big\} \tag{9}$$

Least Squares Temporal Differences

A set of basis functions is used to create a value function approximation. Let $\phi_f(s)$ be a basis function, where $f \in \mathcal{F}$ is a feature. The value function approximation is given by Equation 10 wherein $\theta = (\theta_f)_{f \in \mathcal{F}}$ is a vector of weights (or coefficients) having one coefficient for each basis function.

$$\bar{J}^a(S_t^a|\theta) = \sum_{f \in \mathcal{F}} \theta_f \phi_f(S_t^a) \tag{10}$$

Since the number of features should be much smaller than the state space, es–timating θ is significantly less computationally expensive than calculating the true value function. However, choosing a set of basis functions can be a challenge. Classi–cal linear regression methods can then be used to estimate θ, which is updated using least squares temporal differences.

Least squares temporal differences is a technique for updating the value function approximation for a fixed policy for infinite horizon discounted problems. Temporal differences are the differences between a current estimate of the value of being in a state and the updated value at the following iteration. Powell [21] describes temporal differences as the change in the estimated value of a state over time. Least squares temporal differences achieves this by fitting θ so as to zero the sum of the temporal differences over every iteration [23].

Bellman Error Minimization using Instrumental Variables

The post-decision state, S_t^a, is the state of the system once an action has been taken, but prior to any exogenous information being realized [23]. One effect of using the post-decision state is that the least squares estimators for θ are usually inconsistent. The inconsistency is due to the need to simulate $\phi(S_t^a)$ based on S_{t-1}^a and the dependency between S_{t-1}^a and the error term [23]. To overcome this, Bellman error minimization is accomplished by applying a least-squares methodology to the Bellman error to create updated θ values. Let Φ be the matrices of fixed basis functions. Equation 11 shows the calculation for implementing instrumental variables Bellman error minimization.

$$\hat{\theta} = [(\Phi_{t-1})^T(\Phi_{t-1} - \lambda\Phi_t)]^{-1}(\Phi_{t-1}^T r_t) \tag{11}$$

Least Squares Instrumental Variables Policy Iteration

Using a linear architecture, it is possible to approximate $J(S_t)$ in the post-decision state, $J^a(S_t^a)$, using a column vector of weights, θ, and a column vector with the basis function elements, $\phi(S_t^a)$. In order to calculate the optimal action, Scott *et al.* [23] provide a policy function which is given in Equation 12.

$$A^{\pi}(S_t|\theta) = \underset{a \in A(S_t)}{\arg\max} \left\{ r(S_t, a) + \lambda \theta^T \phi(S_t^a) \right\} \tag{12}$$

To implement IVAPI as shown in Algorithm 1, at each iteration i a random post-decision state $S_{t-1,i}^a$ is generated; the basis function, $\phi(S_{t-1,i}^a)$, is recorded; and the next pre-decision state $S_{t,i}$ is simulated. The optimal action is chosen using the current estimation of θ as indicated by Equation 12, and the resulting contribution and basis function are recorded. This process is then repeated over N iterations. Once the policy evaluation loop is complete, a policy improvement step is executed in which θ is updated using the N observed contributions and basis function evaluations. An instrumental variable method is used for this update to avoid inconsistent least squares estimators for θ. The policy improvement step is executed at each of N iterations [23].

There are four main components to this algorithm: the pre- and post-decision states, random generation of the post-decision state, the basis functions, and the approximate policy iteration methodology. Each of these is discussed.

Pre/post decision state

The pre-decision state space is $\mathcal{S}$, the state space outlined in the MDP formulation. For example, in a 2-COP problem instance, the pre-decision state is $S_t = \{x_1, x_2, v, k\}$ where x_i indicates the current inventory level at COP i, v represents the total number of CUAVs, and k represents the current map. The post-decision state variable has $2+2G$ dimensions where the number of CUAVs deployed to each COP is concatenated

Algorithm 1 Approximate Policy Iteration Algorithm with Instrumental Variables Bellman Error Minimization

```
Initialize $\theta$
for j = 1 to $N$ (Policy Improvement Loop)
    for i = 1 to $M$ (Policy Evaluation Loop)
        Simulate a random post-decision state, $S^a_{t-1,i}$
        Record $\phi(S^a_{t-1,i})$
        Simulate the state transition to get $S_{t,i}$
        Determine the decision, $a = A^\pi(S_{t,i}|\theta)$
        Record $C(S_{t,i}, a)$
        Record $\phi(S^a_{t,i})$, the observation of $E[\phi(S^a_{t,i})|S^a_{t-1,i}]$
    End
    $\hat{\theta} = [(\Phi_{t-1})^T(\Phi_{t-1} - \lambda\Phi_t)]^{-1}(\Phi^T_{t-1}C_t)$ (Policy Improvement)
    Update $\theta$ using generalized harmonic step size rule
End
```

to the pre-decision state to form the post-decision state. For example, a pre-decision state for the 2-COP instance may be $S_t = \{4, 8, 3, 1\}$, while the post-decision state could be $S^a_t = \{4, 8, 1, 1, 0, 2\}$. This example indicates that there are 4 units of inventory at COP 1; 8 units of inventory at COP 2; one CUAV remaining at the BSB; and that the current threat map is 1. The number of vehicles, the third dimension of the state space in this example, is updated in the post-decision state to reflect the number of CUAVs remaining at the BSB and not being sent on a resupply mission. Here, the post-decision state implies that in addition to the information from the pre-decision state, zero CUAVs are deployed to COP 1 and 2 CUAVs are deployed to COP 2. In the post-decision state, no information is known about the outcome of the CUAV resupply missions (SS, SF, or F outcomes).

Random post decision state

In order to randomly generate the post-decision state, the following computations are completed. First, a discrete uniform random variate is generated for v_t, the current number of operational CUAVs, from its state space $\{1, 2, ..., V\}$. Next, a random number of CUAVs to deploy is generated, using a discrete uniform random number between zero and the minimum of the number of crew available and the number of CUAVs, $\{0, 1, ..., min(v_t, \kappa)\}$. Finally, a random number of CUAVs from those deployed is allocated to each COP using a multinomial distribution. This procedure creates the post-decision state for the number of CUAVs available and the actions taken. The post-decision inventory states are generated using a random uniform number between zero and the capacity of the COP, and the random post-decision state for the map is determined using a discrete uniform random number between one and the number of maps available. Alternatively, when we initialize the ADP with inventory at full capacity, we instead use the capacity at each COP. To simulate a transition to the next pre-decision state, a set of multinomial random variates are generated for $Z_t = (Z_{1t}, Z_{2t}, ..., Z_{Gt})$, the outcomes of the resupply missions for all COPs. Transition to the next pre-decision state is based on these outcomes.

Basis functions

Basis functions are chosen based on the minimum number of features necessary to provide an adequate solution. The approximation strategy for determining the value function approximation based on the basis function is shown below:

$$\bar{J}(S_t|\theta) = \sum_{f \in \mathcal{F}} \theta_f \phi_f(S_t) \tag{13}$$

Since we are approximating the post-decision state, we determine the value func– tion approximation based on the post-decision state.

$$\bar{J}^a(S_t^a|\theta) = \sum_{f \in \mathcal{F}} \theta_f \phi_f(S_t^a) \tag{14}$$

Powell [21] notes that choosing the components of the basis function is quite challenging. For a robust approach, potential basis functions need to be throughly explored. In this thesis, basis functions which consist of a first order model or a second order model, as well as functions with indicator variables and interaction terms, are investigated.

API methodology

In order to determine the optimal action to take at the pre-decision state, a deterministic optimization technique is used to solve Equation 12. As shown in Equation 12, this optimal action depends on the chosen contribution function. For small instances of the MILIRP, exhaustive enumeration of all possible actions is completed, and the optimal action is chosen. This technique is acceptable due to the small size of the action space for the small COP instances. For the 12-COP instance, a linear integer program is used to find the optimal action.

Using the approximate policy iteration algorithm with instrumental variables Bellman error minimization, an approximate dynamic program is created for the MILIRP. Using this general formulation of the ADP algorithm we consider specific instances of the MILIRP including the 2-COP, 3-COP, and 12-COP problems.

Table 2. Table of Notation

A^{π}	=	policy function
a	=	action, number of CUAVs to deploy to each COP
B	=	Borel subset
C	=	COP capacity
d	=	daily COP demand
G	=	number of COPs
H	=	known joint probability distribution of Z
I	=	indicator variable
J	=	total expected reward
K	=	number of threat maps
Q	=	CUAV capacity
r	=	reward/contribution function
S	=	state of system
t	=	time epoch
v	=	current number of CUAVs
V	=	number of CUAVs
W	=	Markov transition function
x	=	inventory at a COP
Z	=	set of random variables of the number of possible SS, SF, and F events
$\mathcal{A}$	=	action space
$\mathcal{F}$	=	set of basis function features
$\mathcal{S}$	=	state space
$\mathcal{T}$	=	set of time epochs
τ	=	stock out cost (penalty)
ψ	=	one-way probability a CUAV successfully reaches its destination
ϕ	=	basis function
θ	=	vector of weights
π	=	policy
λ	=	discount factor
κ	=	number of crews
$\triangle$	=	absorbing state
Φ	=	matrix of fixed basis functions
Ω	=	probability of remaining in the low threat map
β	=	probability of remaining in the high threat map

IV. Computational Example

Using the general formulation of the MILIRP presented in Chapter 3, we find the policy for a problem instance with 12 COPs. In order to do this, two additional instances of the MILIRP with two and three COPs are analyzed. For the 2- and 3-COP instances, the exact and myopic solutions are first determined. Basis functions for the ADP are explored to find the ADP's optimal parameters. Finally, an ADP is created for the 12-COP instance.

4.1 MDP Parameterization

The MILIRP is formulated as an infinite horizon MDP where time is discretized into six-hour time periods. This discretization allows the day to be divided into four equal periods. We assume that any single CUAV sortie can be completed during a single period. A single CUAV sortie includes maintenance, fueling, loading, and unloading actions.

We examine a 12-COP instance of the MILIRP with direct delivery. This number is arrived at by considering the maximum dispersal of platoons in a battalion; each COP is manned by a single platoon. With three platoons in a company and four companies in a battalion, a battalion sized force could occupy 12 COPs. We test the ADP at the limits of this feasible region. To aid in creating the 12-COP ADP, both 2- and 3-COP instances are also analyzed.

Each COP has a consumption rate and storage capacity based on the number of personnel at the COP. A General Dynamics report [10] indicates that 8,000 pounds of supplies are consumed by a platoon per day. With four periods in one day, 2,000 pounds (or one ton) of supplies per period are necessary for sustainment. For the remainder of the thesis, we discretize supply units into one-ton units. We conserva–

tively assume that that the COP's capacity is three times the daily demand, bringing COP capacity to 12 tons. We assume that the necessary number of supplies to resupply all COPs are available at the BSB and that the BSB never runs out of supplies. This assumption is reasonable since the BSB is supplied via fixed wing airlift. With undisputed air superiority in Afghanistan, supplies arrive to the BSB with certainty.

CUAV capabilities are increasing as research and development of the systems continue. At present, Lockheed Martin's K-MAX unmanned aircraft system helicopter has successfully transported payloads of three tons at sea level and two tons at 15,000 feet [16]. As recently as 2012, Lockheed Martin announced that the K-MAX routinely transported 4,200 ton load in combat conditions [17]. As a conservative estimate, we use the two-ton capacity of the K-MAX as the CUAVs' capacity in the computational example.

Parameterization of the number of CUAVs available is determined based on the Army's Tactical Unmanned Aircraft System (TUAS) platoon [8]. In September 2010, the TUAS platoon consisted of two crews and four CUAVs [18]. As of 2014, the Army continues to add additional aircraft and crews to the platoon. For this thesis, we parameterize the number of crews at two, and the number of CUAVs at four. The number of crews indicates the number of CUAVs which can be deployed simultaneously. Consider, for example, when one CUAV platoon is supporting the BSB. Out of the four vehicle CUAV fleet, at most two CUAVs can be deployed during a time period.

The ψ values represent the probability a CUAV successfully travels between the BSB and a COP for a specific map. Ideally, an intelligence unit would discretize the AO and assign risk levels to each pixel within the tessellated region. This risk level would take into account threats such as the probability of inclement weather, mechanical issues, and hostile enemy actions. The transition between maps can be created based on seasons or fighting intensity. The least risky path for each map could then be found and used to parameterize the ADP. For the computational example, we choose to explore the case of $K = 2$ threat maps, one with a low threat environment and one with a high threat environment. We create reasonable ψ values with higher values on the low threat map and lower ψ values on the higher threat map. We use values suggested by McCormack [18] for the map transition probabilities.

When a COP's inventory level is depleted, a COP is immediately resupplied via GLOC to full capacity and a penalty is applied. The penalty for a COP needing GLOC resupply would be solicited from a logistics officer in the field and would depend on the difficulty of resupply to the COP via GLOC and the relative importance of using ALOC to resupply the COP. We parameterize the penalty function as ten times the capacity at the COP. This creates a strong enough incentive to ensure the COP is resupplied by CUAV when possible without causing a catastrophic event when GLOC is needed.

We choose a discount factor, λ, that successfully balances future needs with present needs. A commonly used value in this area is 0.98, and it is the value we use herein.

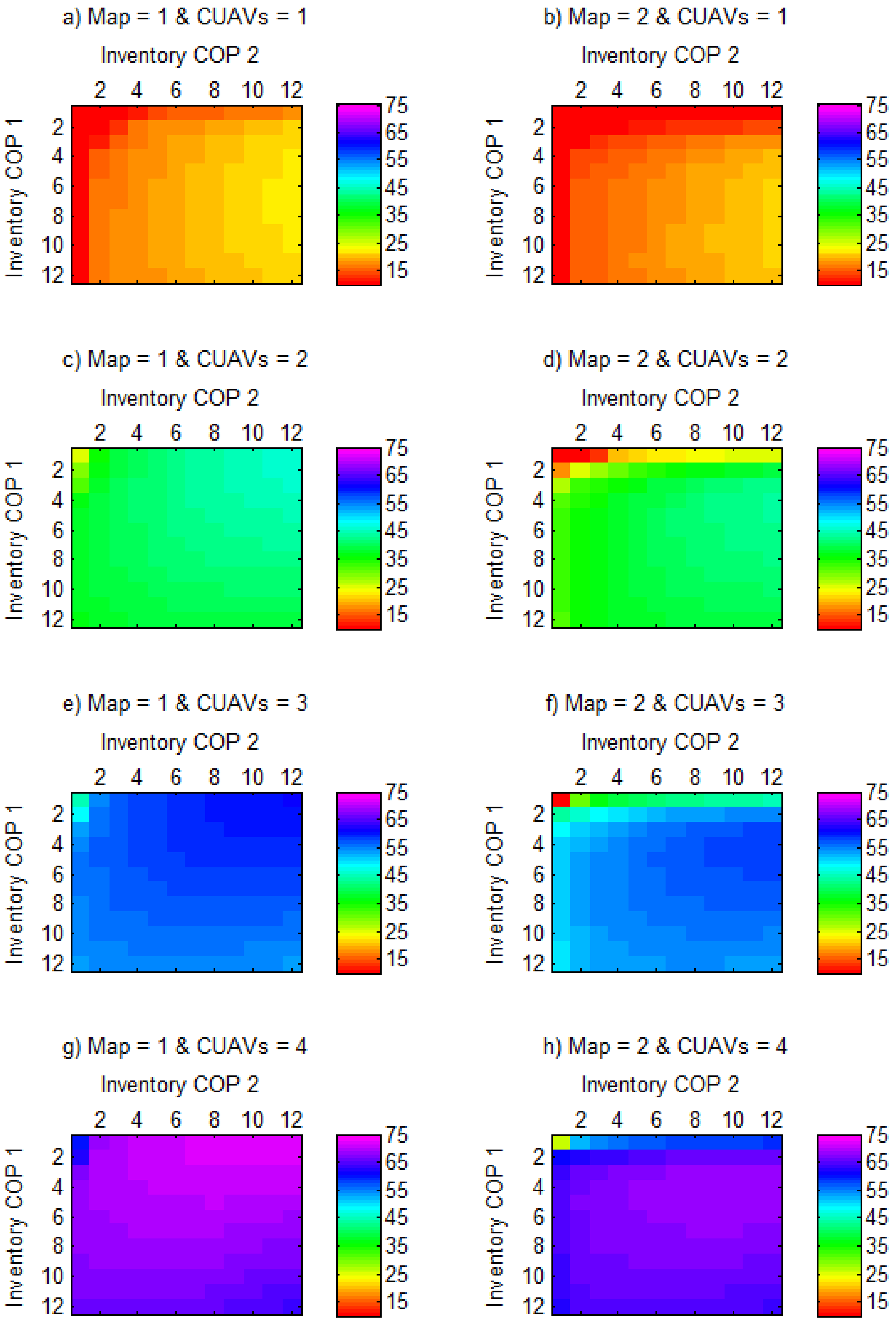
4.2 2-COP Problem Instance

Optimal Policy.

The 2-COP instance of the MILIRP provides initial insight into the problem. We parameterize the ψ_{ik} values as follows: $\psi_{11} = 0.99$, $\psi_{21} = 0.95$, $\psi_{12} = 0.80$, and $\psi_{22} = 0.90$. We assume a single CUAV platoon is present at the BSB. This results in four CUAVs being available and two crews. In this chapter, the optimal resupply policy and value function for the two COP problem instance are presented. With the four dimensions of the inventory level at COP 1, the inventory level at COP 2, the number of vehicles available, and the map, an optimal action can be enumerated for each possible system state.

The results are partitioned by map and number of vehicles available, creating eight categories. In each of these categories a table of optimal policies is presented in Appendix B based on inventory levels. A pair of numbers representing the optimal policy is provided using an (i, j) notation where i indicates the number of CUAVs deployed to COP 1 and j indicates the number of CUAVs deployed to COP 2.

The value function is presented in Figure 1 where the color indicates the value of being in a particular state. Although the legend indicates that the value function fluctuates between 15 and 75, the fluctuation is actually between -119 and 79. Values below 15 are given a red color and values above 75 are given a purple color. This choice was made to create a better visual distinction between the values. The state space is displayed so that Map 1 is on the left and Map 2 is on right, while the available number of CUAVs increases by row.



a) Map = 1 & CUAVs = 1

b) Map = 2 & CUAVs = 1

c) Map = 1 & CUAVs = 2

d) Map = 2 & CUAVs = 2

e) Map = 1 & CUAVs = 3

f) Map = 2 & CUAVs = 3

g) Map = 1 & CUAVs = 4

h) Map = 2 & CUAVs = 4

Figure 1. 2-COP J^* Optimal Policy

Trends are noticeable in the value function. Map 1 results in a higher value function overall than Map 2 (when the number of CUAVs are equal), a result of lower ψ-values in Map 2. As the number of CUAVs increases, the value of being in a particular state increases. Table 3 shows the value of being in each map-vehicle state combination when we fix the inventory state at half capacity for both COPs. The marginal value column of Table 3 indicates the value of one additional CUAV for each map. The marginal value of a CUAV decreases and we observe increased marginal value in Map 1 over Map 2. This indicates that additional CUAVs are more valuable in Map 1.

Table 3. 2 COP: CUAV Marginal Value

# CUAVs	Value: Map 1	Value: Map 2	Marginal Value: Map 1	Marginal Value: Map 2
1	18.98	17.31	-	-
2	42.00	40.02	23.01	22.71
3	57.82	55.98	15.82	15.96
4	69.77	68.08	11.95	12.11

Appendix B provides specific analysis on the optimal policy and optimal value function for each subfigure in Figure 1. There are noticeable trends in the optimal policy over the entire state space. The shape of the optimal policy, as displayed by the tables, depends on the current map when there is more than one CUAV available. This is likely due to the fact that the number of CUAVs that can be deployed is limited by the number of crews (two in this computational example). Overall, CUAVs are less likely to be deployed under Map 2 than Map 1, a result of a lower ψ-value for Map 2. This is depicted by the number of (0,0) actions in the Map 2 optimal policy. Under Map 2, as inventory increases at each COP, a CUAV is less likely to deploy to that COP. However, as the number of CUAVs available increases, the states where CUAVs deploy increases, reflecting the ability to take more risk with an increased number of vehicles. Finally, we observe that, for lower numbers of CUAVs available, the policy of sending one CUAV to each COP is dominated by sending two CUAVs to a single COP. This may be due to the desire to avoid the penalty by ensuring that at least one of the COPs avoids a GLOC resupply.

For every map-CUAV state combination there is a notable decrease in the value function when inventory at COPs 1 and 2 are extremely low. A local maximum is evident under Map 1 when inventory at COP 1 is one and inventory at COP 2 is high, except when the number of CUAVs available is one. Under Map 2, the local maximum occurs when inventory at COP 1 is between three and nine and inventory at COP 2 is twelve. Finally, we observe that the value function is not maximized when both COPs have maximized inventory levels, due to the fact that a reward is not gained for any supplies delivered which exceed the COP's capacity.

Myopic Policy.

The myopic policy selects actions only considering the reward function in the current time period and without consideration of how the decision will affect the future. The policy is arrived at by letting $\lambda = 0$, creating no value from the future outcome of the decision. The myopic policy provides a comparison for the optimal and ADP solutions.

Figure 2 displays the value function for all state combinations when using a myopic policy. Specific analysis of particular optimal policies and optimal value functions are discussed in Appendix B. Overall, the figure shows that Map 2 has slightly lower values of being in a particular state for a given number of CUAVs than Map 1. This can be seen by comparing the graphs on the left and right of the same row. This is due to the higher probability of successful route completion in Map 1. Additionally, looking from top to bottom on the figure, an increase in the value of being in a state is seen. This is due to the increased value in having more CUAVs available. While these trends in the myopic policy are consistent with trends from the optimal policy, there are differences between the myopic and optimal value function representations. When one CUAV is available (subplots a and b), the value of being in a particular state is even lower amongst some of the states in the myopic policy than the optimal policy. In addition, while the value of being in a state for the four CUAV case reaches 75 in the optimal policy, the myopic solution only reaches into the 50s. This reflects the fact that the myopic solution is suboptimal.

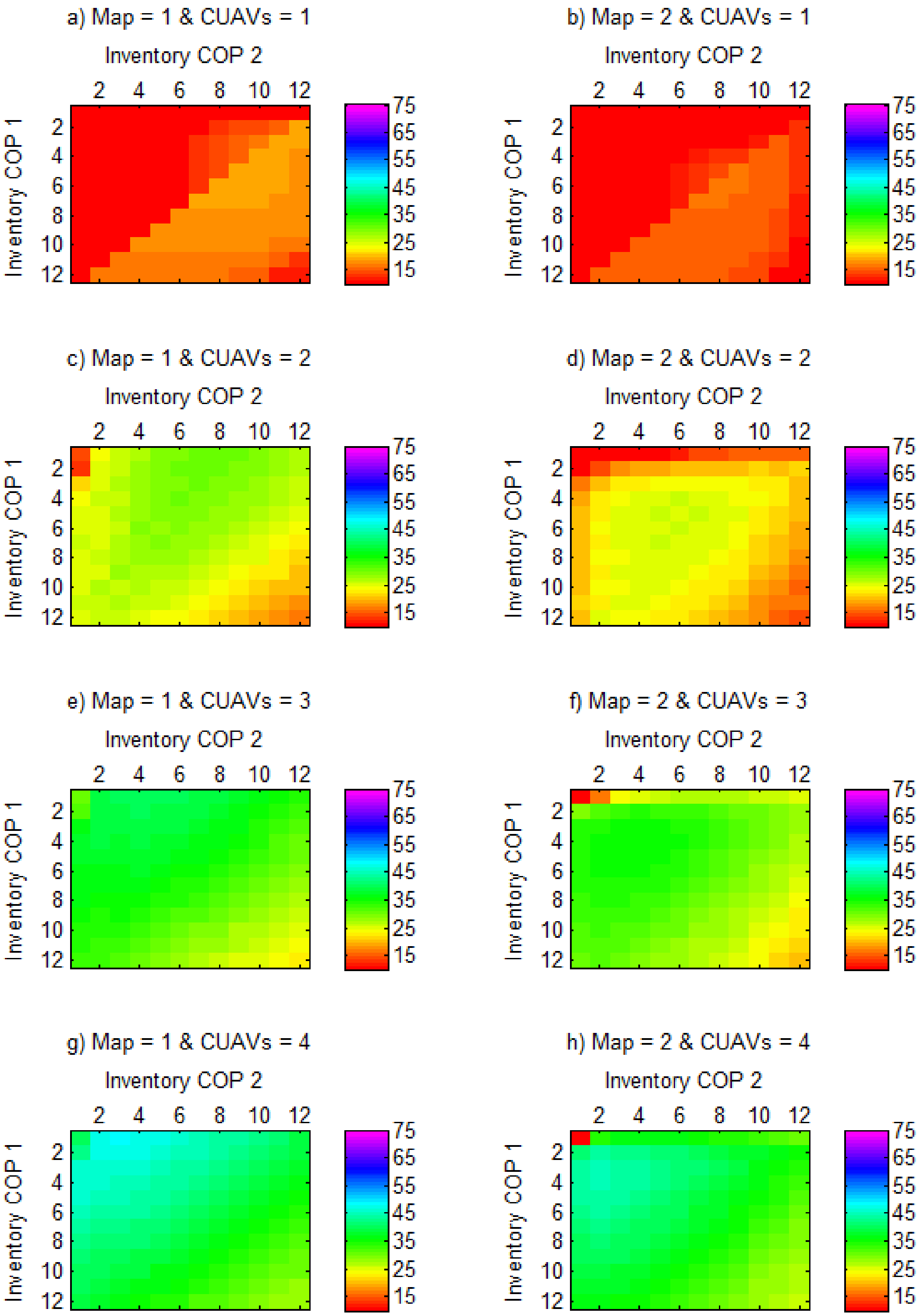


Figure 2. 2-COP Myopic Policy $J*$

ADP Policy.

The ADP policy is created using the least squares approximate policy iteration algorithm with instrumental variables bellman error minimization (IVAPI) algorithm. The challenge in developing this algorithm is creating a set of basis functions that produces adequate results. To explore potential basis function options, first and second order models (both with and without interaction variables) are explored. Indicator variables particular to the problem are also examined. This includes terms which indicate when supplies at the COPs are below a single period's demand, and variables which indicate when there are no CUAVs remaining but the recommended action is to deploy CUAVs. Finally, we also include variables which are constructed by dividing one over the current inventory at each COP. We include an intercept in the basis function to capture the average value of the reward when all other terms in the regression are zero. Once an ADP policy is determined using the IVAPI algorithm, the ADP policy is then evaluated using a policy evaluation algorithm to obtain exact value function results. Moreover, three cases of initialization are reported: both COPs initialized at 50%, 75%, and 100% of capacity. This allows the value functions of the ADP, optimal, and myopic policies to be compared at the three initial system states of interest.

The IVAPI algorithm is implemented using $M = 2000$ for the inner policy evaluation loop and $N = 12$ for the outer policy improvement loop. Figure 3 compares the ADP's performance over fifty replications using the five sets of basis functions.

For each replication of each algorithm implementation, the final policies produced are then evaluated using the exact policy evaluation method. The average percent optimal for each algorithm implementation is shown in Equation 15 where π_i denotes the policy produced at replication i and S_0 is the initialization state.

$$\% \text{ of optimal} = \frac{1}{50} \sum_{i=1}^{50} \frac{\hat{J}^{\pi_i}(S_0)}{J^*(S_0)}. \tag{15}$$

The myopic policy performed the poorest in terms of percent optimality, and its performance only decreased as the inventory initialization percentage increased. The ADP polices performed better than the myopic solution. The second order model with interactions and indicator variables outperformed the other basis functions substantially, reaching over 90% optimality with all three initializations. However, there is a concern that the use of such a large set of basis functions will become computationally expensive as the number of COPs increases.

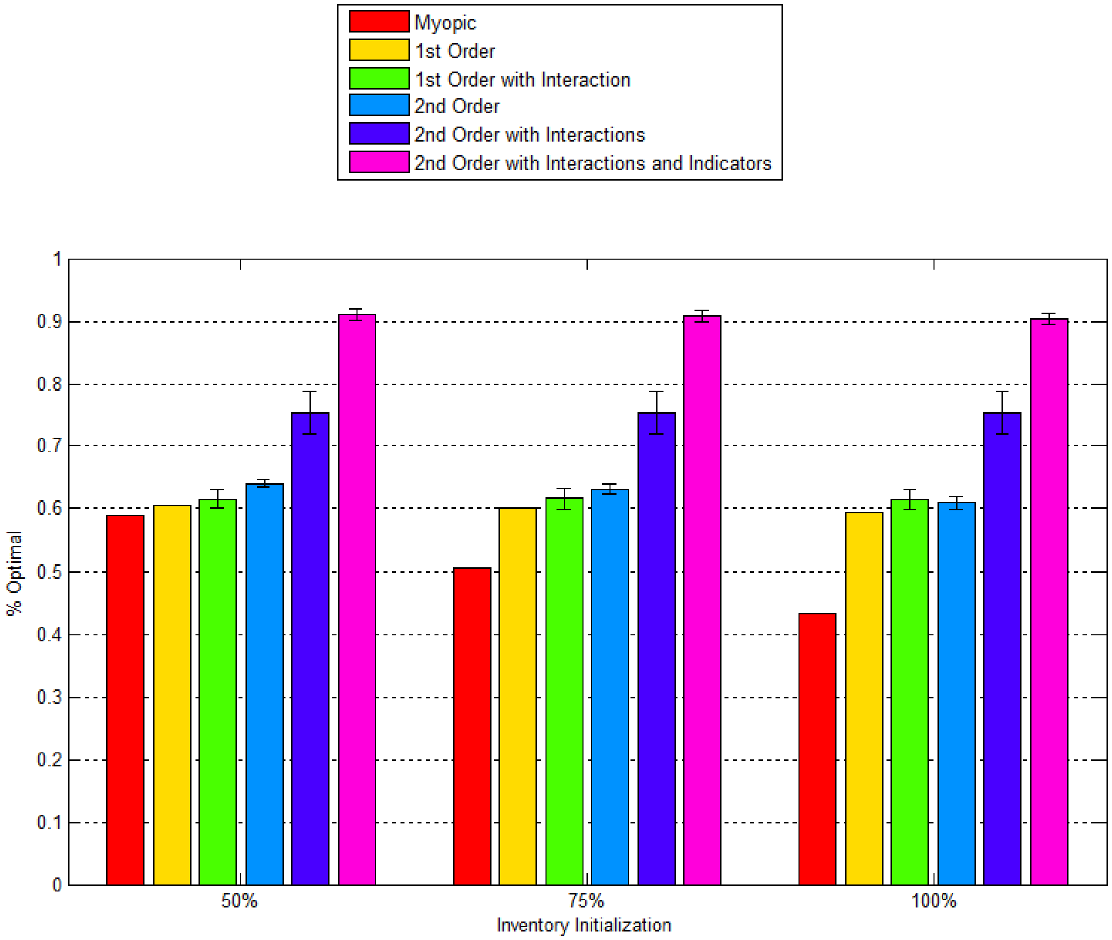


Figure 3. 2-COP Solution Quality with Smoothing

Figure 4 compares the ADP's performance over fifty iterations with the same five basis functions, but excludes the smoothing function. The first order model actually improves when no smoothing is applied whereas the first order with interactions and second order models do not show a statistically significant difference. However, the second order model with interactions and the second order model with interaction and indicator variables models decrease in performance when smoothing is removed.

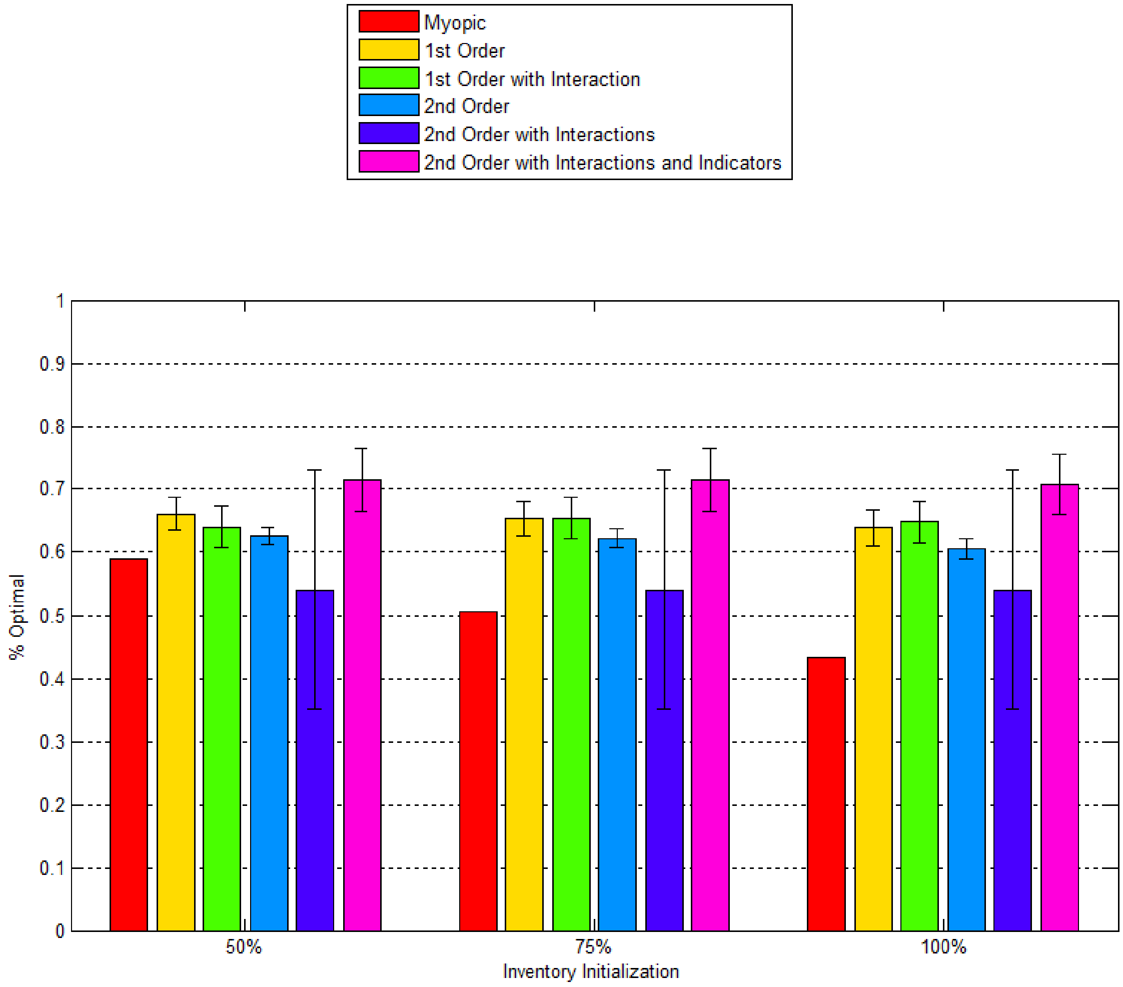


Figure 4. 2-COP Solution Quality without Smoothing

Using the best basis function set identified (second order model with interactions and indicator variables), a simulation is created to observe the performance of the ADP over a thirty day period (i.e. 120 6-hour decision periods). Appendix B contains analysis of these simulations.

4.3 3-COP Problem Instance

The 3-COP instance adds a third COP to the BSB's area of operations. We parameterize the ψ-value between the third COP and the BSB as 92% under both maps. Additionally, we now increase the number of available CUAVs to six and assume a crew of three is available. All other parameters remain the same as in the 2-COP problem instance.

Optimal Policy.

Figure 5 provides the value of the optimal policy for select combinations of the state space. While the legend indicates that the values are between -100 and 50, in reality these values fall as low as -239 and as high as 89, which are represented by dark red and dark blue respectively. Four general trends are observed. First, as we move from left to right on the figure the number of available CUAVs increases, as does the value of being in a particular state. This increase reflects the value an additional available CUAVs adds. Second, moving from the first row to the third row, and from the fourth row to the sixth row, we observe an increase in the values. This increase reflects the increasing value when the inventory at COP 3 increases. Third, the top three rows represent state combinations under Map 1 while the bottom three rows represent state combinations under Map 2. We observe that the value of being in a particular state increases slightly under Map 1 as compared to Map 2. This increase reflects the higher ψ-values for COP 1 and COP 2 in Map 1. Fourth, we observe that, as the inventory at COP 1 or COP 2 (respectively on the horizontal and vertical axes) increases, the value of being in a particular state also increases (with a few exceptions). These trends parallel the conclusions made when examining the

2-COP optimal policy. Appendix C provides analysis on six optimal policies from this graphic for further insight. The conclusion from this analysis indicates that the same general conclusions from the 2-COP example are supported with the 3-COP example.

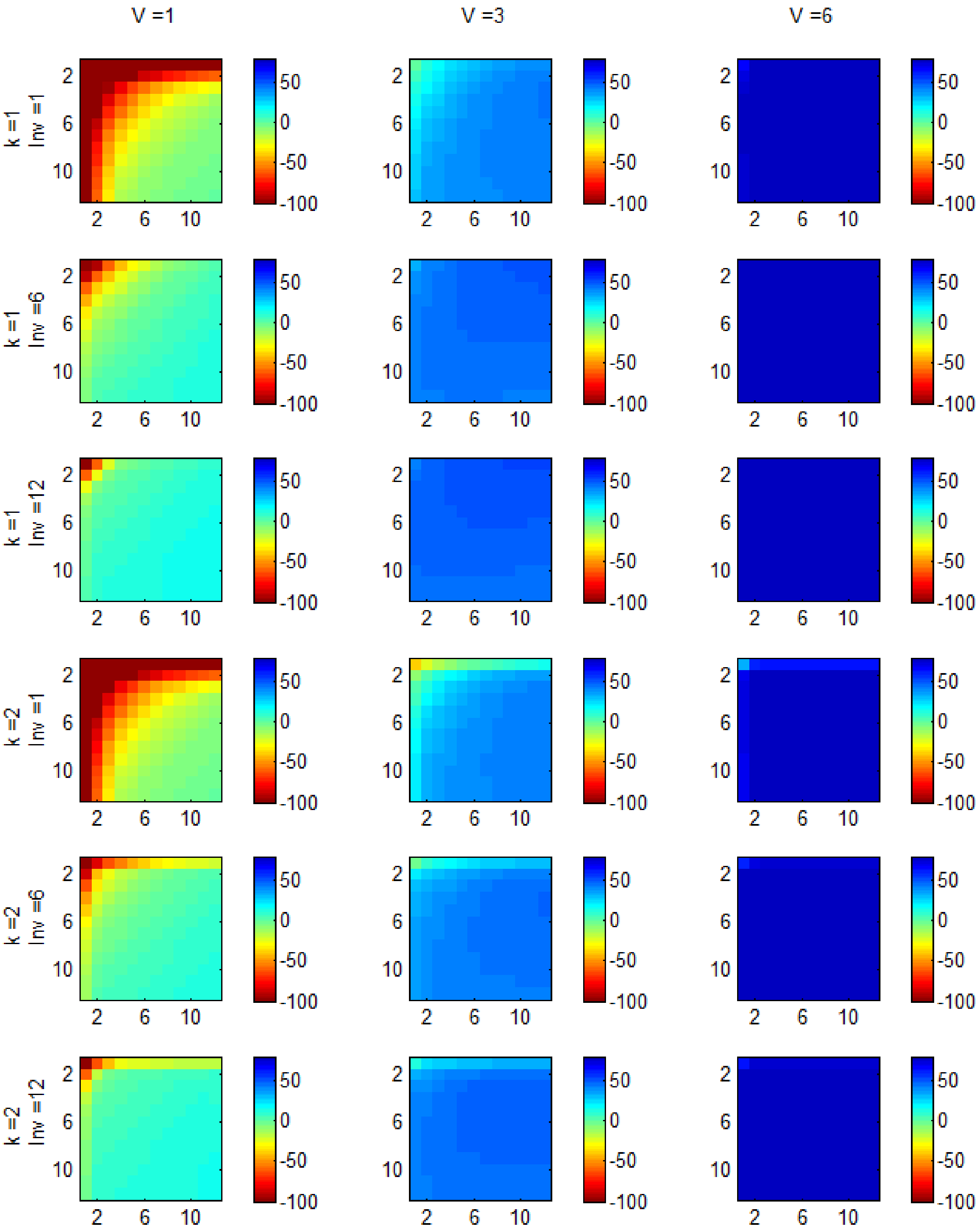


Figure 5. 3-COP $J*$, Optimal Policy

Myopic Policy.

The myopic policy is determined using policy evaluation with $\lambda = 0$ which optimizes the reward function in the current time epoch without consideration of the future. In Appendix C, we consider the optimal policy tables under the same state spaces as the 3-COP optimal solution. This analysis provides the conclusion that in every case for the myopic policy, all available CUAVs (limited by the number of crews) are deployed in a manner that maximizes the reward for the single time epoch. The myopic policy conclusions for the 3-COP problem parallels the conclusions from the 2-COP myopic policy.

ADP Policy.

We create an ADP policy using the instrumental variables approximate policy iteration (IVAPI) algorithm with N = 30 and M = 4000. We test the five basis functions from the 2-COP problem instance. Again, we evaluate an ADP policy using a policy iteration algorithm to obtain exact value function results. We report on the ADP performance for three initialization states: 50%, 75%, and 100% of the COP's capacity. As in the 2-COP instance, this allows the ADP, optimal, and myopic policies to be compared at the three initial system states of interest.

We test the ADP over 100 replications and report the percent optimal for each of the five basis function sets as well as the myopic policy. When calculated with smoothing, the results indicate that the first order basis functions perform the best over all initialization states, as seen in Figure 6. We also test the basis functions without smoothing; the results are shown in Figure 7. Smoothing clearly improves performance over all initialization states.

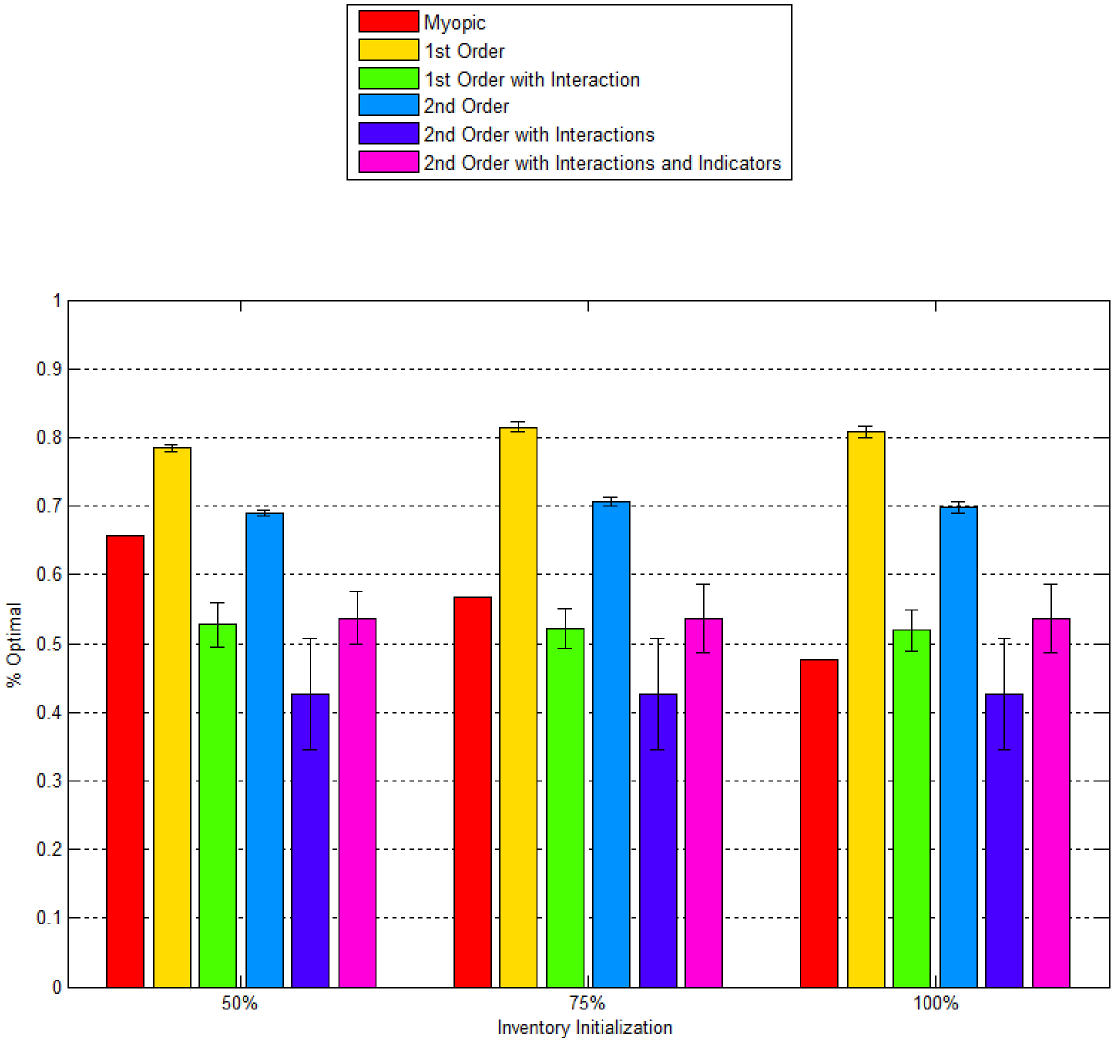


Figure 6. 3-COP Solution Quality with Smoothing

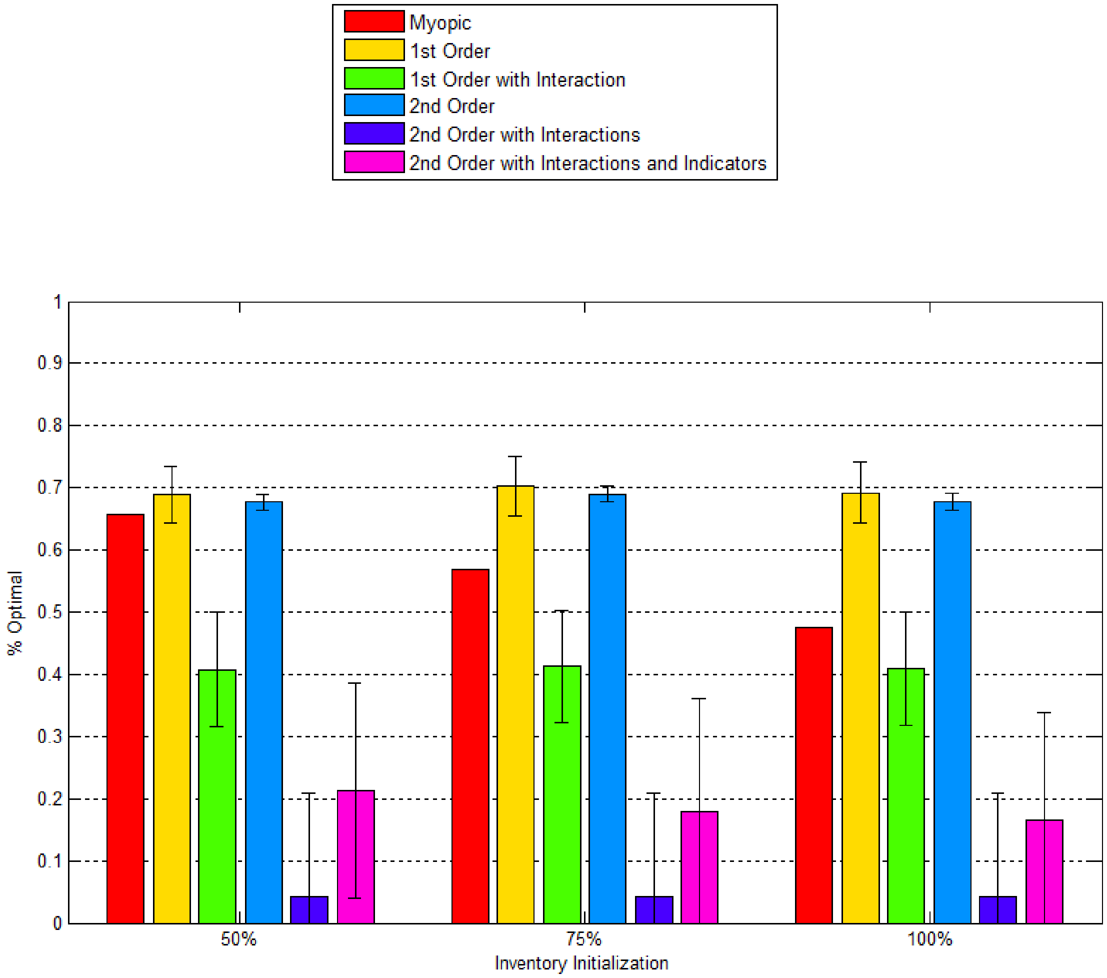


Figure 7. 3-COP Solution Quality without Smoothing

With two small instances of the MILIRP explored, we now investigate the stock–-out penalty in the 2- and 3-COP problems.

4.4 Investigation of the Stock-out Penalty, τ

We explore the penalty value for stock-out, τ. Initially, we parameterize the value of $\tau = -10(C_i)$ which is ten times the capacity at COP i. We investigate three additional parameterizations of the stock-out value: $\tau = \frac{-10(C_i)}{2}$, $\tau = \frac{-10(C_i)}{4}$, and $\tau = 0$. In order to assess the results of changing τ, we simulate optimal, myopic, and ADP policies for each τ value over 100 replications each with a simulation length of

120 periods. We collect data on three statistics, averaged over the 100 replications: the number of GLOC incidents, the number of tons delivered via CUAV, and the period in which the last available CUAV is destroyed. We initialize the simulation with each of the COPs supplied to full capacity.

Tables 4 and 5 display the results of the simulation when τ is changed for the 2- and 3-COP problems. The optimal and approximate policies deliver the most cargo via CUAV when the penalty is set to zero, an average of 138 tons of supplies via CUAV for the 2-COP problem. Additionally, the number of GLOC incidents falls to an average of 7.8 for both the optimal and myopic policies for the 2-COP problem when no penalty is applied. The optimal policy outperforms the myopic policy on average for both the GLOC and ALOC statistics, but only slightly.

Given these results, we decide to remove the penalty and set $\tau = 0$. This removes a difficult-to-parameterize value from the model and simplifies the formulation of the problem. We move to the 12-COP problem instance under the assumption that there is no penalty when stock-out occurs.

Table 4. 2-COP Stock-out Penalty Investigation

2-COP

	Optimal		ADP		Myopic	
	Mean	Stdev	Mean	Stdev	Mean	Stdev
	$\tau = 0$					
1^{st}	-	-	0.61	0.18	-	-
1^{st} with Int.	-	-	0.94	0.06	-	-
2^{nd}	-	-	0.65	0.06	-	-
2^{nd} with Int.	-	-	0.94	0.05	-	-
2^{nd} with Int. and Ind.	-	-	0.94	0.05	-	-
Periods until V = 0	96.73	29.97	96.75	29.41	17.75	8.46
Tons Delivered (Air)	138.95	46.82	138.73	45.94	32.75	15.80
GLOC Incidents	7.83	3.92	7.85	3.87	16.48	1.53
Total Delivered	240	240	240	240	240	240
	τ = .25C					
1^{st}	-	-	0.86	0.00	-	-
1^{st} with Int.	-	-	0.91	0.02	-	-
2^{nd}	-	-	0.84	0.03	-	-
2^{nd} with Int.	-	-	0.87	0.11	-	-
2^{nd} with Int. and Ind.	-	-	0.88	0.03	-	-
Periods until V = 0	64.58	27.07	66.26	29.07	17.75	8.46
Tons Delivered (Air)	115.31	52.47	106.59	53.41	32.75	15.80
GLOC Incidents	9.65	4.48	14.21	6.49	16.48	1.53
Total Delivered	240	240	240	240	240	240
	τ = .5C					
1^{st}	-	-	0.83	0.06	-	-
1^{st} with Int.	-	-	0.82	0.04	-	-
2^{nd}	-	-	0.78	0.07	-	-
2^{nd} with Int.	-	-	0.80	0.09	-	-
2^{nd} with Int. and Ind.	-	-	0.85	0.06	-	-
Periods until V = 0	64.53	26.96	62.64	27.03	17.75	8.46
Tons Delivered (Air)	115.28	52.33	109.41	51.78	32.75	15.80
GLOC Incidents	9.62	4.48	10.99	4.66	16.48	1.53
Total Delivered	240	240	240	240	240	240
	τ = C					
1^{st}	-	-	0.60	0.00	-	-
1^{st} with Int.	-	-	0.62	0.07	-	-
2^{nd}	-	-	0.63	0.03	-	-
2^{nd} with Int.	-	-	0.76	0.15	-	-
2^{nd} with Int. and Ind.	-	-	0.91	0.04	-	-
Periods until V = 0	64.58	27.07	66.26	29.07	17.75	8.46
Tons Delivered (Air)	115.31	52.47	106.59	53.41	32.75	15.80
GLOC Incidents	9.65	4.48	14.21	6.49	16.48	1.53
Total Delivered	240	240	240	240	240	240

Table 5. 3-COP Stock-out Penalty Investigation

3-COP

	Optimal		ADP		Myopic	
	Mean	Stdev	Mean	Stdev	Mean	Stdev
	$\tau = 0$					
1^{st}	-	-	0.74	0.09	-	-
1^{st} with Int.	-	-	0.92	0.01	-	-
2^{nd}	-	-	0.72	0.02	-	-
2^{nd} with Int.	-	-	0.90	0.02	-	-
2^{nd} with Int. and Ind.	-	-	0.91	0.02	-	-
Periods until V = 0	100.24	26.02	89.03	29.08	23.48	12.34
Tons Delivered (Air)	168.24	47.04	160.78	51.68	56.77	26.43
GLOC Incidents	14.91	3.98	16.37	4.32	24.20	2.31
Total Delivered	360	360	360	360	360	360
	τ = .25C					
1^{st}	-	-	0.87	0.00	-	-
1^{st} with Int.	-	-	0.79	0.03	-	-
2^{nd}	-	-	0.87	0.01	-	-
2^{nd} with Int.	-	-	0.73	0.03	-	-
2^{nd} with Int. and Ind.	-	-	0.83	0.02	-	-
Periods until V = 0	46.57	17.00	48.46	16.92	22.73	10.83
Tons Delivered (Air)	119.43	45.76	87.39	33.66	53.62	21.82
GLOC Incidents	18.89	4.03	28.06	4.34	25.14	2.48
Total Delivered	360	360	360	360	360	360
	τ = .5C					
1^{st}	-	-	0.85	0.01	-	-
1^{st} with Int.	-	-	0.74	0.07	-	-
2^{nd}	-	-	0.84	0.02	-	-
2^{nd} with Int.	-	-	0.64	0.09	-	-
2^{nd} with Int. and Ind.	-	-	0.74	0.03	-	-
Periods until V = 0	45.43	17.18	36.79	15.41	22.73	10.83
Tons Delivered (Air)	117.79	47.18	94.12	39.49	53.62	21.82
GLOC Incidents	18.99	4.16	21.91	4.01	25.14	2.48
Total Delivered	360	360	360	360	360	360
	τ = C					
1^{st}	-	-	0.80	0.02	-	-
1^{st} with Int.	-	-	0.52	0.10	-	-
2^{nd}	-	-	0.70	0.02	-	-
2^{nd} with Int.	-	-	0.52	0.32	-	-
2^{nd} with Int. and Ind.	-	-	0.54	0.15	-	-
Periods until V = 0	45.00	16.80	36.79	15.41	22.73	10.83
Tons Delivered (Air)	117.36	45.88	94.12	39.49	53.62	21.82
GLOC Incidents	19.03	4.07	21.91	4.01	25.14	2.48
Total Delivered	360	360	360	360	360	360

4.5 12-COP Problem Instance

We formulate the 12-COP problem as the final computational example wherein 12 COPs represent the maximum dispersal of platoons in a battalion across an area of operations. We parameterize the 12-COP problem instance as follows. We create ψ-values via a continuous uniform distribution that is bounded between 0.8 and 1 for the high threat map and between 0.99 and 1 for the low threat map. This parameterization balances the possibility of failing to make a delivery with providing a realistic risk level at which a commander would deploy a CUAV. We use $K = 2$ threat maps, representing a low and a high threat map. Additionally, given the results from investigating τ, we remove the penalty for stock-out.

With respect to the inner maximization problem in which a best action must be selected for the current decision epoch, complete enumeration is possible for the smaller problem instances. However, for the 12-COP problem instance, complete enumeration is computationally expensive. Instead, we develop an integer program (IP) to obtain a solution for the inner maximization problem. The IP uses the following defined terms.

decision variables:

a_i, number of CUAVs sent from the BSB to resupply COP i

y_i, indicator variable where:

$$y_i = \begin{cases} 1 & \text{if } \psi_i Q a_i + x_i - d_i \leq 0 \\ 0 & \text{if } \psi_i Q a_i + x_i - d_i > 0 \end{cases}$$

constants:

θ_i = coefficient value corresponding to the action taken COP i

$\theta_{vehicles}$ = coefficient value corresponding to the number of CUAVs available

IP:

$$\max \sum_{i=1}^{G} a_i(\psi_i Q + \lambda(\theta_i - \theta_{vehicles})) + \tau_i y_i$$

s.t.:

(1) $\sum_{i=1}^{G} a_i \leq \min(\kappa, v)$, total number of CUAVs constraint

(2) $\psi_i Q a_i + x_i - d_i \leq C_i$, COP's capacity constraint

(3) $\psi_i Q a_i + x_i - d_i \geq -M y_i$, GLOC penalty constraint

(4) $a_i \in \mathbb{Z}^+$.

We simplify the IP with the assumption that $\tau = 0$.

$$\max \sum_{i=1}^{G} a_i(\psi_i Q + \lambda(\theta_i - \theta_{vehicles}))$$

s.t.:

(1) $\sum_{i=1}^{G} a_i \leq \min(\kappa, v)$, total number of CUAVs constraint

(2) $\psi_i Q a_i + x_i - d_i \leq C_i$, COP's capacity constraint

(3) $a_i \in \mathbb{Z}^+$.

We then develop ADP policies using the IP within the IVAPI algorithm. We use first order basis functions for two reasons. First, when testing the 3-COP problem instance over five sets of basis functions, we found that the first order model performed the best. Second, the first order model allows for a linear integer program to be used, rather than a non-linear IP. This significantly simplifies the inner maximization problem.

4.6 Experimental Design

We create set of experiments to assess the proposed ADP's solution quality, computational effort, and robustness [2]. To understand the effect of parameterization on the performance of the ADP, we create a design of experiments. Three response variables are considered: the number of ground resupply incidents (GLOC), the number of tons delivered via CUAV (ALOC), and the number of vehicles that remain at the end of the simulation. It is important to note that the ALOC response variable is reported in total tons while the GLOC response variable is reported in total number of incidents. For each GLOC incident, 12 tons of supplies are delivered. We also record computation times for the ADP to determine the computational effort needed to solve the MILIRP. Finally, we assess the robustness of the algorithm by experimenting with problem factors and algorithmic factors. In order to report these values, a simulation is performed once the ADP policy has been created. We record the three response variables at three different simulation lengths: 1-month, 2-month, and 3-month horizons, simulating over 100 replications per treatment.

Four problem characteristics are investigated: the number of COPs (G), the number of vehicles initially available (V), and the one-step transition matrix for the threat maps. The one-step transition matrix contains two problem factors: Ω and β. We denote the probability of remaining in a low threat map as Ω, and the probability of remaining in a high threat map as β. The probability of transitioning from a low threat map to a high threat map is represented by $1 - \Omega$ while the probability of transitioning from a high threat map to a low threat map is $1 - \beta$.

Each of the four problem factors are considered continuous variables. To determine the high and low factor levels for the number of COPs, we run preliminary experiments. The results indicate that the upper limit of where the ADP policy outperforms the myopic policy in terms of supplies delivered via ALOC is 18 COPs.

When we explore beyond this bound to consider 27 COPs, we find that the myopic policy delivers three times the supplies via ALOC than the ADP policy. Therefore, we use nine as the low level and 15 as the high level for the number of COPs. This parameterization allows the center factor level, 12, to represent the maximum number of platoons in a battalion. For the number of CUAVs, four is used as the low factor level and eight as the high level. This allows the upper bound of the factor to represent two platoons of CUAVs as defined by the Department of the Army [9]. Since CUAV units are organized in a 2:1 ratio of CUAVs to crews, we parameterize the number of crews as half the number of CUAVS initially available. The transition matrix values, Ω and β are explored at the 0.2 and 0.8 levels. The lower bound, 0.2, represents a low probability of returning to the current threat map while the upper bound represents a high probability of returning to the current threat map.

Four algorithmic features are also explored. The number of outer loops (N) and inner loops (M) in the the ADP algorithm are investigated. For the inner loop, M, values between 3,000 and 7,000 are considered. The center value of 5,000 has shown to be adequate for some parameterizations of the 12-COP problem; investigating smaller and larger numbers of loops provides insight into how the performance of the ADP changes with different computational efforts. For the outer loop, N, values between 10 and 30 are used. These bounds are chosen as $N = 30$ has been shown to provide adequate results for the ADP when compared to the myopic policy. By investigating values lower than 30, the performance of the ADP can be assessed for lower computation times. The use of Bellman error minimization alone (L1) or with instrumental variables (L2) is also considered as a two level categorical variable. We denote this factor as IV. Finally, the use of smoothing is also investigated by either applying smoothing (L1), or by not applying smoothing (L2). This final algorithm feature is also a categorical variable and is denoted as SM. The problem and algorithmic factors and their associated levels are shown in Table 6.

Table 6. Factor Settings for Factorial Design

	Description	Factor	Low (-1)	Center (0)	High (1)
Problem Factors	number of COPs	G	9	12	15
	number of CUAVs	V	4	6	8
	probability of remaining in a low threat map	Ω	0.2	0.5	0.8
	probability of remaining in a high threat map	β	0.2	0.5	0.8
Algorithmic Factors	number of inner loops	M	3000	5000	7000
	number of outer loops	N	10	20	30
	instrumental variables	IV	Off (L1)	-	On (L2)
	smoothing	SM	On (L1)	-	Off (L2)

A fractional-factorial design with center runs is implemented. We create a 2^{8-2} resolution V design with a quarter fraction of eight factors in 64 runs. We anticipate that the eight factors previously discussed may have an effect on one of the response variables and are explored at two levels: high (1) and low (-1). We denote the center runs using (0). The resolution V design dictates that some two factor interactions are aliased with three factor interactions. We use an additional four center points (each with one of four combinations of the two categorical variables) to bring the total number of treatment runs to 68. Using this experimental design, we create ADP policies by calculating the θ coefficients for the basis functions. Once this is complete, we use a simulation to obtain the response variable statistics for both the ADP policy and the myopic policy.

Each experiment is conducted in MATLAB R2014b on an Intel(R) Xeon E5-1620 3.6 GHz processor having 32 GB memory. While all experiments are conducted on one type of machine, experiments are performed on different individual computers. When reporting computational effort, only the time for the ADP algorithm to run is recorded; overhead operations and simulation times are not included. We conduct the two simulations per treatment (one after determining an ADP policy, and one utilizing the myopic policy) over 100 replications. We consistently seed the experiments in both the ADP algorithm and the simulation to decrease the variability of the results.

4.7 Results and Analysis

The fractional-factorial design is used to identify the significant factors in the experiment and provide a basis for analysis. Using this design, we estimate all eight single factor terms as well as all 36 two factor interaction terms and some three factor iterations. The results of the experiment for each response variable at the end of the three-month simulation are shown in Table 7. The column of "Coded Factor Levels" shows the pattern of low and high levels for each factor in the treatment, in the order they are shown in Table 6. The computation time necessary to execute the ADP code is provided. Additionally, the mean and standard deviations for the ALOC, GLOC, and number of CUAVs remaining responses are provided for both the ADP and myopic policies. The final column provides a calculation of the difference in the ALOC response variables between the ADP policy and the myopic policy after 360 periods.

The results from the experiment are shown in Table 7; the results from specific runs provide interesting insights. We first examine the results from the experimental run which produced the largest ALOC value over a three month period, Run 27. This run combines the use of instrumental variables and smoothing, the low number of COPs (9), and the high number of CUAVs (8). In this treatment, we observe that after a one month period (not the three month period shown in Table 7), 744 tons of supplies are delivered via ALOC, and 310 tons are delivered via GLOC (29.5 incidents). This ALOC delivery accounts for 70.6% of the total supplies delivered during this period. We then consider the myopic policy's results for this treatment level: 360.3 tons are delivered via ALOC in the first month, and 673.7 tons are delivered via GLOC (56 incidents); only 57% of the total tons are delivered via ALOC. For the same treatment, we observe that the ADP significantly increases the amount of supplies delivered via ALOC. We now consider the results of the designed experiment.

We first consider the ALOC response variable: the number of tons which are de–livered via CUAV over the 360 period simulation. The analysis provides an indication of which factors are statistically significant in affecting the tons of supplies delivered via ALOC. We also check assumptions before proceeding; we verify the equal vari–ance and normality assumptions using the normal probability plot and a plot of the residuals vs. predicted values. The plots confirm that the normality assumption is upheld as well as the constant variance assumption. Plots of the residuals vs. the factor values also confirms that constant variance in the residuals is, for the most part, maintained.

In this analysis, a p-value of ≤ 0.05 is considered significant. Using a screening experiment, we generate a regression model by determining the significant factors in the experiment, which are shown in Table 9. This table indicates that all but one of the main factors are significant in the model. This absence of N as a significant factor indicates that the number of outer loops used in the ADP algorithm does not significantly effect the ALOC response. This is valuable information in terms of computation time as the number of outer loops can be decreased to 10 without significantly affecting the response. The remaining seven main factors exert the most influence on the response variable as they account for 55.7% of the total variance in the model. The two-factor interactions account for 38.4% of the total variance, followed by only two significant three factor interactions that contribute only 1.4% of the variance in the model. The fact that two- and three-factor interactions are significant complicates the interpretation and analysis of the model. It is important

Table 7. Experiment Results, 360-period Horizon

			ADP Policy			Myopic Policy			ALOC
Run	Coded Factor Levels	Comp Time (sec)	ALOC (tons)	# GLOC Incidents	# Vehicles Remaining	ALOC (tons)	# GLOC Incidents	# Vehicles Remaining	Diff. (tons)
1	- - - - - - - +	535.3	73.51 ± 6.56	260.35 ± 0.56	0 ± 0	89.79 ± 7.12	259.8 ± 0.64	0 ± 0	-16.3
2	- - - - - + + - -	1204.0	641.13 ± 22.22*	212.66 ± 1.85	1.89 ± 0.2	89.79 ± 7.12	259.8 ± 0.64	0 ± 0	551.3
3	- - - - + - + -	1549.9	648.11 ± 19.8*	212.07 ± 1.65	1.92 ± 0.21	89.79 ± 7.12	259.8 ± 0.64	0 ± 0	558.3
4	- - - - + + - +	3632.5	81.55 ± 6.48	259.8 ± 0.58	0 ± 0	89.79 ± 7.12	259.8 ± 0.64	0 ± 0	-8.2
5	- - - + - - + -	517.2	91.39 ± 8.21*	259.49 ± 0.77	0 ± 0	51.31 ± 4.16	263 ± 0.37	0 ± 0	40.1
6	- - - + - + - +	1201.7	44.43 ± 4.1	263.13 ± 0.38	0 ± 0	51.31 ± 4.16	263 ± 0.37	0 ± 0	-6.9
7	- - - + + - - +	1545.8	47.03 ± 3.91	262.87 ± 0.38	0 ± 0	51.31 ± 4.16	263 ± 0.37	0 ± 0	-4.3
8	- - - + + + + -	3619.0	98.81 ± 7.97*	258.48 ± 0.7	0 ± 0	51.31 ± 4.16	263 ± 0.37	0 ± 0	47.5
9	- - + - - - + +	518.8	347.61 ± 28.51*	237.31 ± 2.4	0 ± 0	248.09 ± 21.65	246.65 ± 1.81	0 ± 0	99.5
10	- - + - - + - -	1212.8	185.31 ± 15.88	251.3 ± 1.36	0 ± 0	248.09 ± 21.65	246.65 ± 1.81	0 ± 0	-62.8
11	- - + - + - - -	1665.5	143.91 ± 12.72	254.42 ± 1.06	0 ± 0	248.09 ± 21.65	246.65 ± 1.81	0 ± 0	-104.2
12	- - + - + + + +	3927.9	512.49 ± 40.98*	223.58 ± 3.39	0.11 ± 0.05	248.09 ± 21.65	246.65 ± 1.81	0 ± 0	264.4
13	- - + + - - - -	558.1	59.63 ± 6.47	262.49 ± 0.53	0 ± 0	83.03 ± 8.33	260.33 ± 0.71	0 ± 0	-23.4
14	- - + + - + + +	1289.0	604.69 ± 26.64*	215.82 ± 2.22	1.41 ± 0.2	83.03 ± 8.33	260.33 ± 0.71	0 ± 0	521.7
15	- - + + + - + +	1556.2	155.95 ± 15.53*	253.3 ± 1.33	0 ± 0	83.03 ± 8.33	260.33 ± 0.71	0 ± 0	72.9
16	- - + + + + - -	3637.8	53.13 ± 5.56	262.69 ± 0.52	0 ± 0	83.03 ± 8.33	260.33 ± 0.71	0 ± 0	-29.9
17	- + - - - - + +	522.2	272.31 ± 17.36*	243.49 ± 1.45	0 ± 0	152.56 ± 9.72	253.37 ± 0.83	0 ± 0	119.8
18	- + - - - + - -	1228.2	119.77 ± 7.63	256.22 ± 0.65	0 ± 0	152.56 ± 9.72	253.37 ± 0.83	0 ± 0	-32.8
19	- + - - + - - -	1570.4	105.91 ± 6.75	257.19 ± 0.56	0 ± 0	152.56 ± 9.72	253.37 ± 0.83	0 ± 0	-46.7
20	- + - - + + + +	3675.2	581.15 ± 39.61*	217.58 ± 3.29	0 ± 0	152.56 ± 9.72	253.37 ± 0.83	0 ± 0	428.6
21	- + - + - - - -	521.5	65.64 ± 4.11	260.53 ± 0.31	0 ± 0	80.75 ± 5.1	259.16 ± 0.4	0 ± 0	-15.1
22	- + - + - + + +	1224.3	200.07 ± 12.24*	249.62 ± 1.06	0 ± 0	80.75 ± 5.1	259.16 ± 0.4	0 ± 0	119.3
23	- + - + + - + +	1569.5	94.75 ± 5.69*	258.39 ± 0.53	0 ± 0	80.75 ± 5.1	259.16 ± 0.4	0 ± 0	14.0
24	- + - + + + - -	3663.2	62.63 ± 4.09	260.91 ± 0.32	0 ± 0	80.75 ± 5.1	259.16 ± 0.4	0 ± 0	-18.1
25	- + + - - - - +	521.7	296.16 ± 19.9	242.19 ± 1.67	0 ± 0	373.93 ± 24.44	234.95 ± 2.05	0 ± 0	-77.8
26	- + + - - + + -	1216.4	1696.81 ± 62.52*	125 ± 5.15	1.42 ± 0.19	373.93 ± 24.44	234.95 ± 2.05	0 ± 0	1322.9
27	- + + - + - + -	1575.4	1849.93 ± 60.93*	112.9 ± 5.01	2.05 ± 0.22	373.93 ± 24.44	234.95 ± 2.05	0 ± 0	1476.0
28	- + + - + + - +	3664.4	332.75 ± 23.58	238.76 ± 1.98	0 ± 0	373.93 ± 24.44	234.95 ± 2.05	0 ± 0	-41.2
29	- + + + - - + -	522.4	1097.73 ± 47.09*	174.62 ± 3.9	1.59 ± 0.24	134.25 ± 12.56	254.65 ± 1.02	0 ± 0	963.5
30	- + + + - + - +	1218.0	116.91 ± 10.29	257.19 ± 0.87	0 ± 0	134.25 ± 12.56	254.65 ± 1.02	0 ± 0	-17.3
31	- + + + + - - +	1565.5	115.62 ± 10.38	257.05 ± 0.87	0 ± 0	134.25 ± 12.56	254.65 ± 1.02	0 ± 0	-18.6
32	- + + + + + + -	3656.2	1340.15 ± 30.36*	155.05 ± 2.5	3.74 ± 0.29	134.25 ± 12.56	254.65 ± 1.02	0 ± 0	1205.9
33	+ - - - - - - -	585.4	81.93 ± 7.08	438.77 ± 0.74	0 ± 0	161.71 ± 14.18	433.43 ± 1.19	0 ± 0	-79.8
34	+ - - - - + + +	1365.0	120.97 ± 9.62	434.23 ± 0.92	0 ± 0	161.71 ± 14.18	433.43 ± 1.19	0 ± 0	-40.7

Table 8. Experiment Results (Cont.), 360-period Horizon

Run	Coded Factor Levels	Comp Time (sec)	ADP Policy ALOC (tons)	ADP Policy # GLOC Incidents	ADP Policy # Vehicles Remaining	Myopic Policy ALOC (tons)	Myopic Policy # GLOC Incidents	Myopic Policy # Vehicles Remaining	ALOC Diff. (tons)
35	+ - - - + - + +	1736.6	66.15 ± 5.92	439.19 ± 0.66	0 ± 0	161.71 ± 14.18	433.43 ± 1.19	0 ± 0	-95.6
36	+ - - - ++ - -	4062.1	100.57 ± 7.89	437.67 ± 0.71	0 ± 0	161.71 ± 14.18	433.43 ± 1.19	0 ± 0	-61.1
37	+ - - + - - + +	576.2	41.23 ± 3.65	442.59 ± 0.42	0 ± 0	89.05 ± 8.11	439.35 ± 0.72	0 ± 0	-47.8
38	+ - - + - + - -	1360.8	37.15 ± 3.33	443.69 ± 0.33	0 ± 0	89.05 ± 8.11	439.35 ± 0.72	0 ± 0	-51.9
39	+ - - + + - - -	1732.2	42.25 ± 3.57	441.9 ± 0.41	0 ± 0	89.05 ± 8.11	439.35 ± 0.72	0 ± 0	-46.8
40	+ - - + + + + +	4046.9	52.77 ± 4.86	441.93 ± 0.53	0 ± 0	89.05 ± 8.11	439.35 ± 0.72	0 ± 0	-36.3
41	+ - + - - - + -	576.7	201.17 ± 14.3	427.4 ± 1.24	0 ± 0	457.79 ± 38.29	408.52 ± 3.2	0.03 ± 0.03	-256.6
42	+ - + - - ++	1367.0	198.49 ± 17.6	427.94 ± 1.55	0 ± 0	457.79 ± 38.29	408.52 ± 3.2	0.03 ± 0.03	-259.3
43	+ - + - + - - +	1868.9	171.81 ± 14.61	430.71 ± 1.29	0 ± 0	457.79 ± 38.29	408.52 ± 3.2	0.03 ± 0.03	-286.0
44	+ - + - + + + -	4372.1	785.69 ± 51.19*	378.58 ± 4.3	0.54 ± 0.14	457.79 ± 38.29	408.52 ± 3.2	0.03 ± 0.03	327.9
45	+ - + + - - - +	622.1	84.91 ± 8.2	438.12 ± 0.8	0 ± 0	156.65 ± 14.64	433.58 ± 1.27	0 ± 0	-71.7
46	+ - + + - + + -	1355.0	298.89 ± 27.76*	419.32 ± 2.39	0.01 ± 0.02	156.65 ± 14.64	433.58 ± 1.27	0 ± 0	142.2
47	+ - + + + - + -	1731.0	138.05 ± 11.56	433.84 ± 1.07	0 ± 0	156.65 ± 14.64	433.58 ± 1.27	0 ± 0	-18.6
48	+ - + + + + - +	4058.2	74.73 ± 7.52	438.91 ± 0.75	0 ± 0	156.65 ± 14.64	433.58 ± 1.27	0 ± 0	-81.9
49	++ - - - - + -	586.5	188.99 ± 10.36	428.5 ± 0.92	0 ± 0	255.48 ± 16.45	423.73 ± 1.34	0 ± 0	-66.5
50	+ + - - - + - +	1368.2	168.9 ± 10.23	430.75 ± 0.92	0 ± 0	255.48 ± 16.45	423.73 ± 1.34	0 ± 0	-86.6
51	+ + - - + - - +	1747.8	130.53 ± 8.06	434.21 ± 0.7	0 ± 0	255.48 ± 16.45	423.73 ± 1.34	0 ± 0	-125.0
52	+ + - - + + + -	4081.0	685.16 ± 46.06	386.81 ± 3.87	0.01 ± 0.02	255.48 ± 16.45	423.73 ± 1.34	0 ± 0	429.7
53	+ + - + - - - +	588.0	89.17 ± 5.54	437.01 ± 0.51	0 ± 0	127.78 ± 8.86	434.7 ± 0.79	0 ± 0	-38.6
54	+ + - + - + + -	1364.0	192.7 ± 11.64*	428.83 ± 1.07	0 ± 0	127.78 ± 8.86	434.7 ± 0.79	0 ± 0	64.9
55	+ + - + + - + -	1762.3	89.64 ± 5.56	436.97 ± 0.54	0 ± 0	127.78 ± 8.86	434.7 ± 0.79	0 ± 0	-38.1
56	+ + - + + + - +	4089.4	83.86 ± 4.99	438.44 ± 0.47	0 ± 0	127.78 ± 8.86	434.7 ± 0.79	0 ± 0	-43.9
57	+ + + - - - - -	592.8	287.29 ± 16.51	421.51 ± 1.42	0 ± 0	676.5 ± 43.38	388.26 ± 3.63	0.03 ± 0.03	-389.2
58	+ + + - - + + +	1361.9	556.69 ± 30.65	397.5 ± 2.6	0 ± 0	676.5 ± 43.38	388.26 ± 3.63	0.03 ± 0.03	-119.8
59	+ + + - + - + +	1753.3	292.56 ± 18.94	419.43 ± 1.62	0 ± 0	676.5 ± 43.38	388.26 ± 3.63	0.03 ± 0.03	-383.9
60	+ + + - + + - -	4096.5	296.72 ± 19.44	420.68 ± 1.66	0 ± 0	676.5 ± 43.38	388.26 ± 3.63	0.03 ± 0.03	-379.8
61	+ + + + - - + +	583.6	121.49 ± 11.34	435.28 ± 0.96	0 ± 0	221.48 ± 18.07	426.75 ± 1.56	0 ± 0	-100.0
62	+ + + + - + - -	1360.1	126.25 ± 9.78	434.42 ± 0.84	0 ± 0	221.48 ± 18.07	426.75 ± 1.56	0 ± 0	-95.2
63	+ + + + + - - -	1771.9	101.52 ± 8.67	436.44 ± 0.72	0 ± 0	221.48 ± 18.07	426.75 ± 1.56	0 ± 0	-120.0
64	+ + + + + + + +	4099.5	192.75 ± 15.77	428.13 ± 1.36	0 ± 0	221.48 ± 18.07	426.75 ± 1.56	0 ± 0	-28.7
65	0 0 0 0 0 0 L1 L1	1816.9	87.81 ± 6.41	348.46 ± 0.58	0 ± 0	145.42 ± 10.03	343.54 ± 0.84	0 ± 0	-57.6
66	0 0 0 0 0 0 L1 L2	1819.4	122.93 ± 9.32	345.5 ± 0.8	0 ± 0	145.42 ± 10.03	343.54 ± 0.84	0 ± 0	-22.5
67	0 0 0 0 0 0 L2 L1	1824.9	867.18 ± 37.37*	282.73 ± 3.11	1.29 ± 0.2	145.42 ± 10.03	343.54 ± 0.84	0 ± 0	721.8
68	0 0 0 0 0 0 L2 L2	1816.4	143.31 ± 11.86	343.86 ± 1.02	0 ± 0	145.42 ± 10.03	343.54 ± 0.84	0 ± 0	-2.1

to note that, when the estimates for main effects and interactions are discussed, all other significant factors are held constant at their average. We use this screening design to develop a regression model with the significant factors from Table 9.

Table 9. Factor Influence on ALOC Response

Source	Sum of Squares	Prob >F	% Contribution
IV	1,804,801.70	<.0001	0.198
Ω	864,342.10	<.0001	0.095
β	601,826.90	<.0001	0.066
SM	579,681.00	<.0001	0.064
G	569,194.80	<.0001	0.063
V	501,830.60	<.0001	0.055
M	145,408.80	0.0003	0.016
IV $\cdot$ Ω	362,584.60	<.0001	0.040
IV $\cdot$ β	146,823.10	0.0003	0.016
IV $\cdot$ SM	686,053.40	<.0001	0.075
Ω $\cdot$ SM	178,844.40	<.0001	0.020
β $\cdot$ SM	64,986.80	0.0116	0.007
IV $\cdot$ G	636,086.00	<.0001	0.070
Ω $\cdot$ G	240,541.20	<.0001	0.026
SM $\cdot$ G	157,569.30	0.0002	0.017
IV $\cdot$ V	205,707.60	<.0001	0.023
Ω $\cdot$ V	242,901.10	<.0001	0.027
SM $\cdot$ V	217,622.30	<.0001	0.024
G $\cdot$ V	186,235.40	<.0001	0.020
IV $\cdot$ M	112,007.40	0.0012	0.012
β $\cdot$ N	55,025.40	0.0196	0.006
IV $\cdot$ β $\cdot$ SM	68,081.90	0.0099	0.007
Ω $\cdot$ V $\cdot$ M	55,519.10	0.0191	0.006

Table 10 provides the significant parameter estimates and their associated p-val–ues. The first term, instrumental variables, indicates that when instrumental variables is not used, the average ALOC response decreases by 163 tons. Thus, using instru–mental variables is important in the formulation of the ADP. The second term, Ω, indicates that a 0.3 increase in the value of Ω increases the ALOC response by 116.2 tons. This means that when the probability of staying in a low threat map increases, the ALOC response variable also increases. This makes sense as the simulation will remain in a low threat map for longer amounts of time allowing more CUAVs to make successful deliveries. The third significant term, IV·SM, indicates that the interaction of smoothing and instrumental variables is also important. This trend can also be observed by considering the four center runs in Table 7. All other factors held constant at their mid-points, out of the four combinations of the IV and SM levels, the combination which results in the highest ALOC response is the combination of instrumental variables and smoothing.

The fourth term captures the interaction between using instrumental variables and the number of COPs. Looking to the seventh term, G, we note that increasing the number of COPs decreases the response variable. This parallels the results from initial testing done on the number of COPs, which indicated that the ADP does not perform as well at higher number of COPs. However, looking back to the fourth parameter (IV·G), we see that when instrumental variables is not utilized, an additional three COPs actually increases the ALOC response variable by 99.7 tons. This conclusion is an example of the difficulty in interpreting models with significant main effects and interactions. Moreover, the positive coefficient for IV·G-term indicates that instrumental variables seems less effective as the number of COPs increases. The fifth factor, β, indicates that increasing the probability of remaining in a high threat map decreases the ALOC response variable by 97 tons. This makes sense, as remaining in the high threat map is more risky and results in fewer CUAVs being deployed and fewer successful deliveries. The next significant factor, smoothing, indicates that applying smoothing significantly increases the ALOC response. The estimate for the final main factor, V, indicates that increasing the number of vehicles by two increases the value of the response by 88.6 tons. This also makes sense as more available CUAVs allows for more potential deliveries.

The next ten significant terms from Table 10 are two-factor interactions. Of note is the fact that, when we consider interactions with G (terms 11, 14, and 16), all the estimates for the terms are negative. This indicates that even a large number of CUAVs, a high probability of staying in the low threat map, or smoothing cannot overcome the negative effect of having a large number of COPs. The 18^{th} term, M, is the lowest significant main effect. Increasing the number of inner loops results in an increasing ALOC value. This makes sense as the higher number of inner loops should allow for a better solution to be determined. Interestingly, the interaction term of N and M is not significant in the model. This is unexpected; we would expect that the interaction between the policy evaluation loop and the policy improvement loop would affect determination of the θ-values and therefore the quality of the solution. The lack of significance for this interaction term may indicate that the factor levels were not spread far enough apart in the experiment to observe a significant difference in the response. Terms 20 and 22 introduce the two significant three-factor interac–tions. It should be noted that Ω·V·M is aliased with another three-factor interaction, IV·β·N. We choose Ω·V·M as the significant factor because N is not found to be a significant factor in the model. Since additional experimentation is not performed within the scope of this thesis, it is not possible to verify this choice. The occur–rence of significant three-factor interactions suggests that interactions between the variables beyond the two-factor interactions are important. Specifically, the IV·β·SM (Term 20) indicates that the combination of smoothing, instrumental variables, and probability of remaining in a low threat map are important in combination. The only occurrence of N in the model is found in the 23^{rd} term which captures the two factor interaction of β·N. With a p-value of 0.019, this term is significant, but it is the least significant of those terms remaining in the model.

Table 10. Coefficient Estimates for ALOC Response

#	Term	Estimate	Lower 95%	Upper 95%	Prob>$\|t\|$
1	IV[L1]	-162.9	-186.6	-139.3	<.0001
2	Ω	116.2	91.8	140.6	<.0001
3	IV[L1]·SM[L1]	-100.4	-124.1	-76.8	<.0001
4	IV[L1]·G	99.7	75.3	124.1	<.0001
5	β	-97.0	-121.4	-72.6	<.0001
6	SM[L1]	92.3	68.7	116.0	<.0001
7	G	-94.3	-118.7	-69.9	<.0001
8	V	88.6	64.2	112.9	<.0001
9	IV[L1]·Ω	-75.3	-99.7	-50.9	<.0001
10	Ω·V	61.6	37.2	86.0	<.0001
11	Ω·G	-61.3	-85.7	-36.9	<.0001
12	SM[L1]·V	58.3	33.9	82.7	<.0001
13	IV[L1]·V	-56.7	-81.1	-32.3	<.0001
14	G·V	-53.9	-78.3	-29.6	<.0001
15	Ω·SM[L1]	52.9	28.5	77.3	<.0001
16	SM[L1]·G	-49.6	-74.0	-25.2	0.0002
17	IV[L1]·β	47.9	23.5	72.3	0.0003
18	M	47.7	23.3	72.1	0.0003
19	IV[L1]·M	-41.8	-66.2	-17.4	0.0012
20	IV[L1]·β·SM[L1]	32.6	8.2	57.0	0.0099
21	β·SM[L1]	-31.9	-56.3	-7.5	0.0116
22	Ω·V·M	-29.5	-53.8	-5.1	0.0191
23	β·N	-29.3	-53.7	-4.9	0.0196

The R^2 value indicates that the model explains 95.5% of the variance in the ALOC response variable. A R^2_{adj} value of 0.931 and the small difference between this value and R^2 indicates that the parameters of the model are well chosen.

Although we are interested in the total amount of cargo delivered via ALOC, we compare the ALOC responses for the ADP policy to the responses for the myopic policy to assess the performance of the ADP algorithm. We use a one-sided t-test to determine whether the difference in the response variable between the myopic policy and the ADP policy are significant at the 0.05 level for each treatment after a three-month period simulation. The null and alternative hypotheses are shown below.

$$H_0 : \mu_{\text{ADP, ALOC}} - \mu_{\text{Myopic, ALOC}} = 0 \; H_1 : \mu_{\text{ADP, ALOC}} - \mu_{\text{Myopic, ALOC}} > 0$$

The results of the t-tests are shown in Table 7. Occurrences where the ADP outperforms the myopic solution by a statistically significant margin are recognized with an asterisk. By examining the results, a pattern is observed. Out of the 68 experimental runs, only 21 result in the ADP policy significantly outperforming the myopic policy for the ALOC response variable, about 31%. However, if we consider only experimental runs which use smoothing and instrumental variables, this percentage increases to 76% with 13 of the 17 values showing a significantly better response. Moreover, if we also only consider experiments done with the low or center number of COPs as a factor setting, the percentage increases to 100% for all nine experiments. This indicates that there are factors which significantly effect the performance of the ADP compared to the myopic. We also note that the ADP seems to do well compared to the myopic when the number of vehicles remaining at the end of the three month simulation is greater than zero. This positive value indicates that the ADP policy has used its CUAV in such a manner that it reserves some CUAVs for future use.

To explore these factors, we create an additional response variable which captures the difference between the myopic and ADP policies by finding the difference between the ALOC values over a three month simulation. These differences are provided in Table 7. Large positive differences indicate that the ADP significantly outperformed the myopic policy. Small differences indicate that the ADP and myopic policies performed similarly. Large negative differences indicate that the myopic policy significantly outperformed the ADP policy.

Using this response variable further, insight into the algorithm features can be gained by considering Figure 8. We notice that the response variable, the difference between the ADP and myopic policies' ALOC values, is affected by the two categorical variables. The combination of using instrumental variables with smoothing produces the greatest positive differences in the policies.

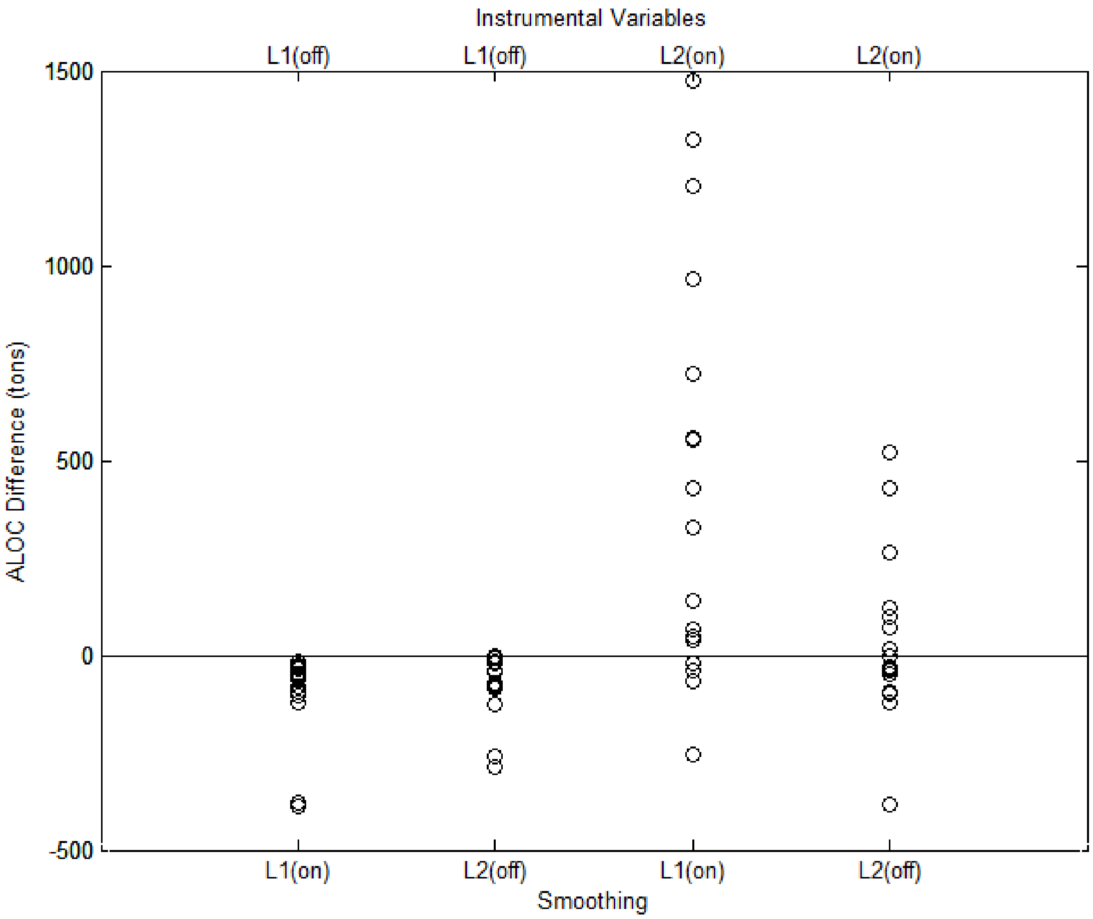


Figure 8. Smoothing and IV

We then create a regression model using the experimental design and the ALOC difference as the regression variable to provide insight into the factors which affect the ADP performance compared to the myopic policy's performance. Table 11 shows the percent contribution for each of the statistically significant terms in the model. We observe that the six main effects contribute to 47.7% of the total variance in the model. Instrumental variables and the number of COPs each contribute significantly, indicating that both terms contribute towards creating a difference between the ADP and myopic solutions. We also recognize that two other algorithmic features, smooth–ing and the number of inner loops, also contribute significantly. Interestingly, neither N nor β contribute significantly in this model in their main effect form. Fifteen two-–

factor interactions are found to be significant in the model. Of note, every two-factor interaction containing the instrumental variables factor is found to be significant with the exception of N. The two-factor interactions account for 45.2% of the total vari– ance in the model, nearly as much as the main effects. This indicates that higher order interactions amongst the variables are important in creating a large difference between the ADP and myopic policies. One three-factor interaction, IV·SM·β, is also found to be significant, although it contributes to less than 1% of the variance in the model.

Table 11. Factor Influence on ALOC Difference Response

Source	Sum of Squares	Prob >F	% Contri
IV	1,804,737	<.0001	0.207
G	1,490,475	<.0001	0.171
SM	579,718	<.0001	0.067
M	145,390	0.0005	0.017
V	133,773	0.0009	0.015
Ω	66,603	0.0153	0.008
IV·G	636,086	<.0001	0.073
IV·SM	740,322	<.0001	0.085
G·SM	157,609	0.0003	0.018
IV·M	111,991	0.002	0.013
IV·V	205,753	<.0001	0.024
G·V	254,949	<.0001	0.029
SM*V	217,669	<.0001	0.025
IV·Ω	362,555	<.0001	0.042
G·Ω	486,088	<.0001	0.056
SM·Ω	178,823	0.0002	0.021
V·Ω	140,550	0.0007	0.016
IV·β	146,881	0.0005	0.017
G·β	125,972	0.0012	0.014
SM·β	65,000	0.0165	0.007
Ω·β	110,656	0.002	0.013
IV·SM·β	83,897	0.0069	0.010

Using the terms found to be significant, we create a regression model. Table 12 provides the significant terms, their coefficient estimates, and the p-value for their t-statistic. Instrumental variables is found to be the most significant factor in the model, which mirrors the result found for the ALOC response variable. However, the importance of the second term, G, is not the same as in the ALOC response model and indicates that on average, there is a 152-ton negative effect on the difference in ALOC values when three additional COPs are considered. The IV·SM interaction term is again found to be significant which supports the conclusions made from Figure 8. Overall, the terms and their order of significance based on p-values are similar for the ALOC response variable and the ALOC difference response variable. Of note is the fact that, for the ALOC difference model, Ω's coefficient is about a fourth the size of the ALOC model, indicating that the effect of Ω is not as important in improving the ADP over the myopic. Additionally, the β term which was in the ALOC model is not statistically significant in the differences model. Instead, it seems that the algorithmic factors as well as the number of COPs are more important in this model. Using the R^2 value, it is understood that 94.6% of the variance in the response variable is captured with the model. Additionally, the R^2_{adj} value of 92.0% indicates that the model is well fit.

Table 12. Coefficient Estimates for ALOC Difference Response

#	Term	Estimate	Lower 95%	Upper 95%	Prob>$\|t\|$
1	IV[L1]	-162.9	-187.9	-137.9	<.0001
2	G	-152.6	-178.4	-126.8	<.0001
3	IV[L1]·SM[L1]	-104.3	-129.3	-79.4	<.0001
4	IV[L1]·G	99.7	73.9	125.5	<.0001
5	SM[L1]	92.3	67.3	117.3	<.0001
6	G·Ω	-87.2	-112.9	-61.4	<.0001
7	IV[L1]·Ω	-75.3	-101.0	-49.5	<.0001
8	G·V	-63.1	-88.9	-37.4	<.0001
9	SM[L1]·V	58.3	32.6	84.1	<.0001
10	IV[L1]·V	-56.7	-82.5	-30.9	<.0001
11	SM[L1]·Ω	52.9	27.1	78.6	0.0002
12	G·SM[L1]	-49.6	-75.4	-23.9	0.0003
13	IV[L1]·β	47.9	22.1	73.7	0.0005
14	M	47.7	21.9	73.4	0.0005
15	V·Ω	46.9	21.1	72.6	0.0007
16	V	45.7	20.0	71.5	0.0009
17	G·β	44.4	18.6	70.1	0.0012
18	IV[L1]·M	-41.8	-67.6	-16.1	0.0021
19	Ω·β	41.6	15.8	67.3	0.0022
20	IV[L1]·SM[L1]·β	36.2	10.4	62.0	0.0069
21	Ω	32.3	6.5	58.0	0.0153
22	SM[L1]·β	-31.9	-57.6	-6.1	0.0165
23	β·N	-29.3	-53.7	-4.9	0.0196

Finally, we further investigate θ, the vector of weights for the basis function, for two particular treatments from the experiment to develop further insight into the ADP. We first consider the θ coefficients resulting from experimental Run 27, which produces the largest ALOC response. By analyzing the θ-values for this particular run, insights into why the ADP performed well are gained. First, by simply graphing the θ-values, it is evident that there is a cutoff between values near zero and those that are not. Values that are near zero indicate that the basis function corresponding to that value did not produce a change in the total discounted reward. For example, the θ-value which corresponds with the current number of vehicles remaining has a value of 48.65 for this particular experimental treatment. This means that, for each additional CUAV, there is an average increase in the total discounted reward of 48.65 tons with all the other variables held constant. Only 20 of the 166 basis functions have corresponding θ-values above 1 or below -1, which we graph in Figure 9. These 20 θ-values fall into four categories of term types: the intercept, number of vehicles, actions, or map-action interactions. The basis function which captures the current number of vehicles has a value of 48 and was discussed above. The basis function coefficients corresponding to the action taken at each COP have values between 43.49 and 49.71. This indicates that deploying an additional CUAV to a particular COP increases the total discounted reward by 43 to 50 tons. Finally, the θ-values for the interactions between the current map and the action taken at each COP varies between -3.33 and -17.9. Due to the fact that the current map is modeled as a binary variable (where the low risk map is zero, and the high risk map is one), deploying CUAVs when in the high risk map decreases the total expected reward between 3.33 and 17.9 tons, depending on the COP. This shows that the basis function is capturing the long-term effect of sending out CUAVs in the more risky map and potentially losing the CUAV.

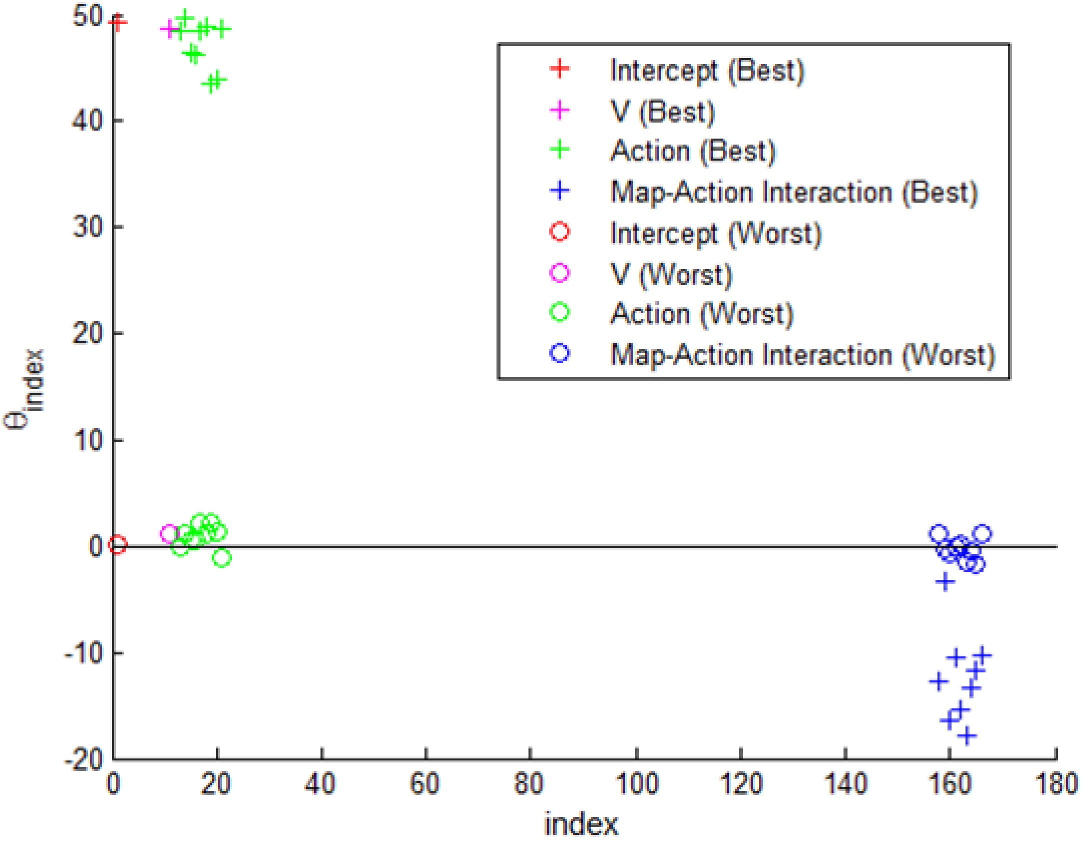


Figure 9. θ Values for Best and Worst Results

We then consider a second set of θ-values determined by choosing the experimental treatment which performed the worst, Run 6. Of note is the fact that, unlike the treatment considered for the best θ-values, this run did not include instrumental variables or smoothing, algorithmic factors which have been found to be important to the performance of the ADP. Looking at the θ-values obtained for this experimental run, it is clear that the magnitude of the values is much smaller than the previous θ-values, varying between -1.77 and 2.26. The poor performance of this particular treatment makes sense as the basis function is not producing θ-values which capture how the reward function changes. As the θ-values for the ADP approach zero, the ADP policy approaches the myopic policy. For example, in the best case run we observed a θ-value of 48.65 for the current number of vehicles. For this worst case run, we obtain a value of 1.22 for the same parameter. Despite the small magnitudes of the θ-values for the worst run, the largest θ-values in magnitude include the same three groupings of basis functions: the number of vehicles remaining, the current action taken at each COP, and the map-action interactions. This indicates that despite the fact that the magnitudes of the θ-values are low, the significant contributors to the total discounted reward remain the same.

V. Conclusions and Recommendations

5.1 Conclusions

First and foremost, this thesis provides the Army and Air Force with insight into the emerging field of unmanned tactical airlift and, specifically, cargo unmanned aerial vehicles. With high casualty rates from ground resupply efforts, the Army looks to unmanned aerial resupply vehicles as a resource which could be used to supplement ground resupply efforts. Every ground convoy *not conducted* provides an opportunity to potentially save lives. The use of CUAVs provides other benefits: the higher flight ceiling and better flight performance in adverse weather conditions makes unmanned helicopters less susceptible to MANPADs, provides for greater maneuverability, and it allows sorties to be scheduled in riskier environments than their manned counterparts. Additionally, a more dedicated platform may enable a more reliable, quicker, and more flexible resupply effort. The addition of a dedicated CUAV unit would also free manned rotary assets for combat missions. However, the CUAVs' key ability is the potential to save lives by partially alleviating the need for ground convoy resupply efforts.

We look to the K-MAX as a specific testament to the Army, Navy, and Air Force's interest in unmanned tactical airlift platforms. After a $45.8 million dollar contract, Lockheed Martin and Kaman Aerospace Corporation successfully deployed three op–tionally manned K-MAX helicopters to Afghanistan in 2012 [15]. During their de–ployment, the K-MAX helicopters were used by Marines in a tactical airlift role to decrease the number of ground convoys necessary, especially in hazardous areas. The Washington Post [15] reported in June of 2014 that "the Marines raved about [the K-–MAXs] utility and dependability" despite one of the three helicopters crashing (with no injuries). Over the duration of the deployment, the Washington Post reported that over 4.5 million pounds of supplies were delivered via the unmanned K-MAX over thousands of sorties.

The K-MAX's performance between 2011 and 2014 laid the groundwork for unmanned tactical airlift to become a reality in today's warfare. With this operational implementation, a capability gap exists regarding how to best apply these assets in a war-time environment. This thesis sought to fill this gap by informing the development of tactics, techniques, and procedures for optimal utilization of CUAV resources for commanders in the field. Proper utilization of CUAVs will prolong the lifespan of the CUAV and increase its utility. By providing procedures for sustaining units via CUAV, we provide decision makers with a potentially lifesaving tool. Although no combat environment will perfectly match the computational example provided in this thesis, decision makers could create their own threat maps and inputs to gain an understanding of a near-optimal policy for deployment of their CUAV resources. Even if the policy is not followed exactly, it will provide a framework for understanding how the CUAVs should be deployed and their expected lifespan, allowing these commanders to better predict their tactical airlift capabilities and needs.

We can also look to the K-MAX's broadening operational role to identify potential areas where this research can inform development of CUAV capabilities. For example, Flightglobal [24] reported on the K-MAX's ability to transport and deploy the Army's unmanned ground vehicle, the Squad Mission Support System, in July 2014. Lockheed Martin has also demonstrated K-MAX's autonomous firefighting capability. Given the funding and interest in developing CUAVs, it is likely that the mission of the CUAV will continue to broaden; it will be important to look to studies like this which can further inform the development and design of the CUAVs. By providing sensitivity analysis on CUAV capabilities such as the number of crew required or the cargo capacity of the CUAV, the value of CUAV capabilities can be examined.

Moreover, analysis results can be used to inform funding and requirement decisions.

The IVAPI algorithmic developed to solve the MILIRP with direct delivery provides a policy for the allocation of CUAV assets to resupply a battalion-sized Army unit. The ADP policy was shown to be successful in outperforming the myopic policy. Experimentation on algorithmic features allowed for the conclusion that the ADP policy improves when high numbers of inner loops are utilized with instrumental variables and smoothing. In terms of problem features, the ADP's performance decreases when a large number of COPs is involved, but the algorithmic is robust to changes in other problem features. Specific combinations of inputs resulted in up to 71% of supplies being delivered via ALOC over a one-month horizon, 65% over a two-month horizon, and 57% over a three-month horizon.

5.2 Limitations

The current IVAPI algorithmic does not perform well when 18 or more COPs are considered. This inability to successfully scale the algorithm to a larger number of COPs could be due to the fact that the basis functions that were chosen do not adequately capture the problem nuances for such large instances. The goal of this research was originally to provide a policy for a 36-COP problem instance, the maximum number of platoons in a brigade sized unit. Instead, we are only able to model a problem instance one-third the size. Additionally, we only experimented on four problem features and four algorithmic features in the designed experiment. Sensitivity analysis on many other factors would be of interest to commanders using this analysis and for informing the development of the ADP. For example, additional analysis on the ψ-values for the high and low threat maps, additional threat maps, COP capacity, the CUAV capacity, the discount factor, and the number of crews could provide further insights.

5.3 Future Research

There are numerous areas for future research on the MILIRP. In terms of for–mulating the problem, the addition of supply classes would bring the model closer to accurately representing the Army's real world resupply procedures. Additionally, a more realistic GLOC resupply decision point (e.g. resupplying when a COP is at half capacity rather than completely depleted) would also significantly increase the utility. Additional accuracy in representing the Army's true procedures could also be gained by modifying how demand is modeled. In this thesis, demand is modeled deterministically, but realistically the demand at a COPs is stochastic. By modeling demand in a stochastic nature a more realistic problem would be modeled. Finally, we explore only a single algorithm for determining an ADP policy; exploration of alternative ADP algorithms may produce results which scale better than the results gained from the IVAPI algorithm.

Appendix A. Acronyms

ADP= approximate dynamic programming

AO = area of operations

BSB = brigade supply battalion

COP = combat outpost

CUAV = cargo unmanned aerial vehicle

F = CUAV does not successfully deliver supplies

GLOC = ground lines of communication

IBCT = infantry brigade combat team

IED= improvised explosive device

IRP = inventory routing problem

IV = instrumental variables

MANPADS = man portable air defense system

MDP = Markov decision process

MILIRP= military inventory routing problem

SF = CUAV delivers supplies to COP, but does not successfully return

SIRP = stochastic inventory routing problem

SM = smoothing

SS = CUAV completes both legs of the journey

TUAS = tactical unmanned aerial system

UAV = unmanned aerial vehicles

VMIP = vendor managed inventory practices

VRP = vehicle routing problem

VRPSD = vehicle routing problem with stochastic demand

Appendix B. Computational Example: 2-COP

Optimal Policy

Table 13 presents the optimal policy when one CUAV is available under Map 1 for each inventory combination. From the optimal policy it appears that at every inventory level combination, a vehicle should be sent either to COP 1 or COP 2. Overall, when inventory is low at one COP and high at the other COP, the depleted COP is sent a CUAV. However, when inventory is above three units at both COPs, the overall policy is to send a CUAV to COP 1. This can be explained by observing that the ψ-value at COP 1 (0.99) is greater than at COP 2 (0.95) for Map 1. This also explains the policy to deploy a CUAV to COP 1 when inventory at both COPs is one. The optimal policy maximizes the expected total discounted reward by deploying the CUAV to the COP with the higher ψ-value. An exception to this is seen at the bottom of the policy chart where a CUAV is sent to COP 2 when inventory at COP 1 is maximized. Another exception appears to be when COP 2 is low in inventory. The system seeks to avoid the penalty, as shown by the optimal action to send a CUAV to COP 2 when inventory at COP 2 is low.

Figure 1a shows the value of being in each combination of inventory states under Map 1 when one CUAV is available. The deep red in the top left corner indicates that the value of being in a state where both COPs have low inventory levels is quite low. The red and orange bars along the top and left of the graph similarly depict the low value of low inventory states. A local maximum when inventory at COP 1 is seven and inventory at COP 2 is twelve is shown in bright yellow. The local maximum's location is likely due to the fact that COP 2's inventory is maximized, while COP 1's inventory is not. Since the optimal policy in this state is to send a CUAV to COP 1, a full reward is gained with 99% certainty.

The optimal policy when a single CUAV is available changes from Map 1 to Map

Table 13. 2-COP Optimal Policy, Map = 1, Vehicles = 1

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	2	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	3	(0,1)	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	4	(0,1)	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	5	(0,1)	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	6	(0,1)	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	7	(0,1)	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	8	(0,1)	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	9	(0,1)	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	10	(0,1)	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	11	(0,1)	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	12	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)

2, as shown in Table 14. First, unlike the optimal policy under Map 1, there are instances where the best action is to not deploy any CUAVs. This is likely due to the fact that the ψ-values are lower at both COPs for Map 2. Given the higher probability of losing a CUAV, the optimal policy indicates that it is better to wait for Map 1 to deploy the CUAVs unless inventory is low. Second, the optimal action when both COPs have inventories of one changes. In Map 2, the optimal action is to send a CUAV to COP 2. This change is due to the higher ψ-value for COP 2 than COP 1 in Map 2.

Figure 1b shows the value of being in a particular state (under Map 2 and when one CUAV is available) and is similar to Figure 1a. Overall, the increased amount of red and orange indicates an overall decrease in the value at each of the inventory states. This decrease is due to the lower ψ-values in Map 2 than in Map 1, as well as the decreased willingness to deploy CUAVs which in turn does not allow for rewards to be gained. However, the dark red in the left top corner and along the left and top row indicates that low inventory states are indicative of low values in those states. The local maximum shifts down to where inventory at COP 1 is 8 or 9 and inventory at COP 2 is maximized at 12.

Table 14. 2-COP Optimal Policy, Map = 2, Vehicles = 1

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	2	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	3	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	4	(0,1)	(0,1)	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	5	(0,1)	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	6	(0,1)	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	7	(0,1)	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	8	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	9	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	10	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	11	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	12	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)

Table 15 presents the optimal policy when two CUAVs are available under Map 1 for each inventory combination. With two CUAVs available, the optimal action can include six possible solutions: (2,0), (0,2), (1,0), (0,1), (1,1), and (0,0). As with the previous optimal policies, when inventory is low at one COP and high at the other, two CUAVs are deployed to the COP needing resupply. When both COPs' inventory levels are low, one CUAV is sent to each COP. There is also a channel in the middle left of the optimal policy where (1,1) is the optimal action. Alternative (2,0) dominates the optimal policy at all levels of inventory for COP 1, and COP 2 levels of inventory between one and nine. This reflects the high ψ-value for Map 1 between the BSB and COP 1. When both COPs have high levels of inventory, either one CUAV is deployed, or none are sent at all, as shown in the bottom right corner of the optimal policy table.

Like the previous two figures, Figure 1c shows a sharp decrease in the value of being in state spaces with low inventory levels for both COPs. There also seems to be a local maximum when COP 1's inventory is low and COP 2's inventory is high, as shown in the top right of the figure. From Table 15 we know the optimal policy in these states is to send two CUAVs to COP 1. This action maximizes the expected reward, resulting in high values of being in these states without the concern that the delivered inventory will exceed capacity. The value of being in a state decreases to a constant level of about 35 over the remaining states. The change from red to green between Figures 1a and 1c show an increase in the value of being in a state when more CUAVs are available.

Table 15. 2-COP Optimal Policy, Map = 1, Vehicles = 2

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	2	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	3	(0,2)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	4	(0,2)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	5	(0,2)	(0,2)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	6	(0,2)	(0,2)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	7	(0,2)	(0,2)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	8	(0,2)	(0,2)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	9	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	10	(0,2)	(0,2)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,0)
	11	(0,2)	(0,2)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,0)
	12	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,1)	(0,1)	(0,0)

The optimal policy changes substantially between Map 1 and Map 2 when two CUAVs are available. Table 16 indicates that the optimal policy is dominated by (0,0), again reflecting the lower ψ-values for Map 2 compared to Map 1. Disparate inventory levels at the COPs results in two CUAVs being deployed to the COP in need of resupply. There are more instances where two CUAVs are deployed to COP 2 than COP 1 due to the increased probability of a successful mission to COP 2. Only when inventory is at one unit for both COPs is a CUAV deployed to both COPs. One exception, when inventory is three at COP 1 and four at COP 2 results in an optimal policy of sending one CUAV to COP 2.

Figure 1d depicts the value function for each inventory level when there are two CUAVs available under Map 2. This figure looks similar to Figure 1c when two CUAVs are available under Map 1. Again, there is a sharp drop off of the value at low inventory levels and a local maximum when COP 1's inventory is four and COP 2's inventory is 12. Overall, the value function seems to decrease on this figure compared to the figure from Map 1, reflecting the lower ψ-values in Map 2.

Table 16. 2-COP Optimal Policy, Map = 2, Vehicles = 2

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	2	(0,2)	(0,2)	(0,2)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	3	(0,2)	(0,2)	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	4	(0,2)	(0,2)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	5	(0,2)	(0,2)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	6	(0,2)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	7	(0,2)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	8	(0,2)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	9	(0,2)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	10	(0,2)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	11	(0,2)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	12	(0,2)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)

The optimal policy with three CUAVs available under Map 1 closely resembles the optimal policy for two CUAVs available under Map 1. This similarity is likely due to the limiting factor of crews. Looking to Table 17, we see that the (1,1) action is optimal over a wider range of inventory levels than in the two CUAV scenario. This difference in policy is due to having a CUAV in reserve.

As with the previous value function figures, the value of being in a particular state increases overall when a CUAV is added, as seen in Figure 1e. Low levels of inventory at both COPs causes a large decreases in the value of being in that state. A local maximum is still depicted at low levels of COP 1 inventory and high levels of COP 2 inventory. The value of being in a particular state decreases slightly as the inventory levels move away from this maximum.

Table 17. 2-COP Optimal Policy, Map = 1, Vehicles = 3

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	2	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	3	(0,2)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	4	(0,2)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	5	(0,2)	(1,1)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	6	(0,2)	(1,1)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	7	(0,2)	(0,2)	(1,1)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	8	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	9	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	10	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,0)
	11	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,0)
	12	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,1)	(0,1)	(0,0)

With three CUAVs available under Map 2 as shown in Table 18, the optimal policy is similar to the optimal policy when two CUAVs are available under Map 2. The difference lies in the expansion of instances where (0,1) is optimal. In the two CUAV scenario, only one (0,1) action was recommended, whereas in the three CUAV optimal policy, (0,1) is recommended for inventory levels at COP 2 between two and four, and all COP 1 inventory levels. The (0,1)'s in general replace (0,0)'s, which is a response to the extra CUAVs available. With three CUAVs, more risk can be taken.

Figure 1f appears similar to Figure 1e. However, there is an overall decrease in the value between these two figures. As with other value function figures with Map 2, there is a local maximum when COP 1's inventory is five and COP 2's inventory is 12.

Table 18. 2-COP Optimal Policy, Map = 2, Vehicles = 3

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	2	(0,2)	(0,2)	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	3	(0,2)	(0,2)	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	4	(0,2)	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	5	(0,2)	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	6	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	7	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	8	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	9	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	10	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	11	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	12	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)

Again, the optimal policy when four CUAVs are available under Map 1 is similar to the two and three CUAV optimal policies under Map 1 as seen in Table 19. The central band of (1,1)'s are slightly changed, and one additional anomaly is observed where a (1,1) replaces a (2,0) and a (0,2) replaces a (1,1).

In Figure 1g we note an overall increase in value function from the figures depicting instances with three CUAVs available. The sharp drop in value at the low inventory levels and the local maximum are again observed.

Table 19. 2-COP Optimal Policy, Map = 1, Vehicles = 4

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	2	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	3	(0,2)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	4	(0,2)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	5	(0,2)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	6	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	7	(0,2)	(0,2)	(1,1)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	8	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	9	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(1,1)	(2,0)
	10	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,0)
	11	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(1,1)	(1,1)	(0,2)	(1,1)	(1,1)	(1,0)
	12	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,1)	(0,1)	(0,0)

Finally, with four available CUAVs under Map 2, the optimal policy closely re– sembles the optimal policy when two or three CUAVs are available under Map 2 (Table 20. In this policy, some (0,1)'s shift to the right when inventory is low at both COPs. This indicates that more risk can be assumed (sending two CUAVs instead of one and sending one CUAV instead of two) with the extra CUAVs available. The need to reserve CUAVs is relaxed.

Figure 1h confirms a slight decrease in value function when switching from Map 1 to Map 2. Additionally, the sharp decrease at at low inventory levels is still seen, as well as the local maximum.

Table 20. 2-COP Optimal Policy, Map = 2, Vehicles = 4

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	2	(0,2)	(0,2)	(0,2)	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	3	(0,2)	(0,2)	(0,2)	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	4	(0,2)	(0,2)	(0,1)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	5	(0,2)	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	6	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	7	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	8	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	9	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	10	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	11	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	12	(0,2)	(0,1)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)

Myopic Policy

Tables 21 and 22 show the myopic policy under Map 1. The myopic policy is to send a CUAV to COP 1 unless inventory at COP 1 is 12 or inventory at COP 2 is one, with two exceptions. This reflects the highest ψ-value under Map 1 occurring between COP 1 and the BSB. The myopic policy remains constant when more than vehicle is available, as shown in Table 22. Except when inventory at COP 2 is one or when inventory at COP 1 is high, the myopic action is to send two CUAVs to COP 1 (again reflecting a high ψ-value under Map 1 for COP 1).

Table 21. 2-COP Myopic Policy, Map = 1, Vehicles = 1

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	2	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	3	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	4	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	5	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	6	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	7	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	8	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	9	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	10	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	11	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	12	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)

Table 22. 2-COP Myopic Policy, Map = 1, Vehicles = 2/3/4

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	2	(0,2)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	3	(0,2)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	4	(0,2)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	5	(0,2)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	6	(0,2)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	7	(0,2)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	8	(0,2)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	9	(0,2)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	10	(0,2)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(2,0)
	11	(0,2)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)
	12	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)

Tables 23 and 24 provide the myopic policy under Map 2. Note that the myopic policy tables for Map 2 are the transposes of the myopic policy tables for Map 1; this is a reflection of the higher ψ-value for COP 2 than COP 1 under Map 2. In general, the maximum number of CUAVs available (limited by crews) is sent to COP 2.

Table 23. 2-COP Myopic Policy, Map = 2, Vehicles = 1

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(0,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
	2	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)
	3	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)
	4	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)
	5	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)
	6	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)
	7	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)
	8	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)
	9	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)
	10	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)
	11	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(1,0)
	12	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)

Table 24. 2-COP Myopic Policy, Map = 2, Vehicles = 2/3/4

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,1)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
	2	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)
	3	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)
	4	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)
	5	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)
	6	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)
	7	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)
	8	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)
	9	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)
	10	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(2,0)
	11	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)	(1,1)
	12	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(1,1)	(1,1)

ADP Policy

Figures 10, 11, and 12 show the inventory level at COPs 1 and 2 as well as the number of CUAVs available over a 30-day period using the optimal policy, a sample ADP policy, and the myopic policy. Figure 10, which was created using the optimal policy, indicates that for this sample path, only two CUAVs were destroyed over the 30-day period. While COP 1 has more stable and higher inventory levels, COP 2's inventory never drops below four units. Although the current map is not displayed in the graphic, it should be noted that it is likely that CUAVs were usually deployed when the system was in Map 1.

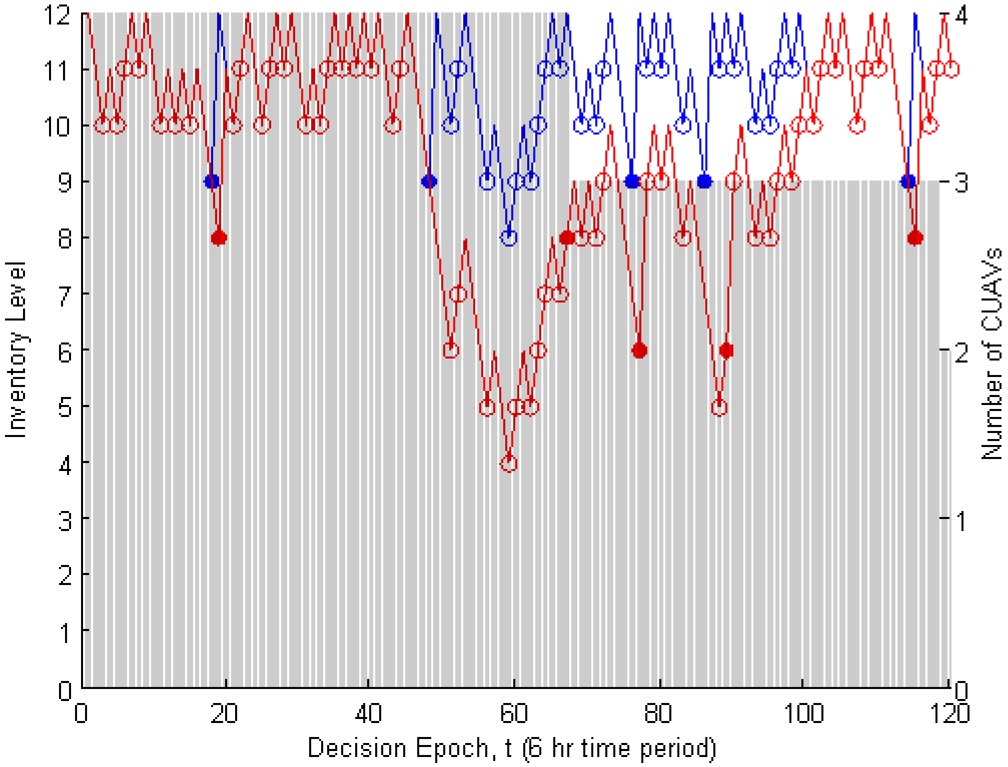


Figure 10. 2-COP 30-day sample path of the optimal policy

Figure 11 shows the sample path when an ADP policy is used in the simulation. After 30 days only two CUAVs are destroyed, and both COPs have inventory levels of 11. Compared to the optimal policy, there is more variation in inventory levels for both COPs. Additionally, sending two CUAVs to a single COP is a more frequent action in the ADP simulation than in the optimal simulation. Again, COP 1 has higher inventory levels in general over the length of the simulation, a product of the higher ψ-values for COP 1 in Map 1 coupled with the tendency to wait for Map 1 to deploy CUAVs.

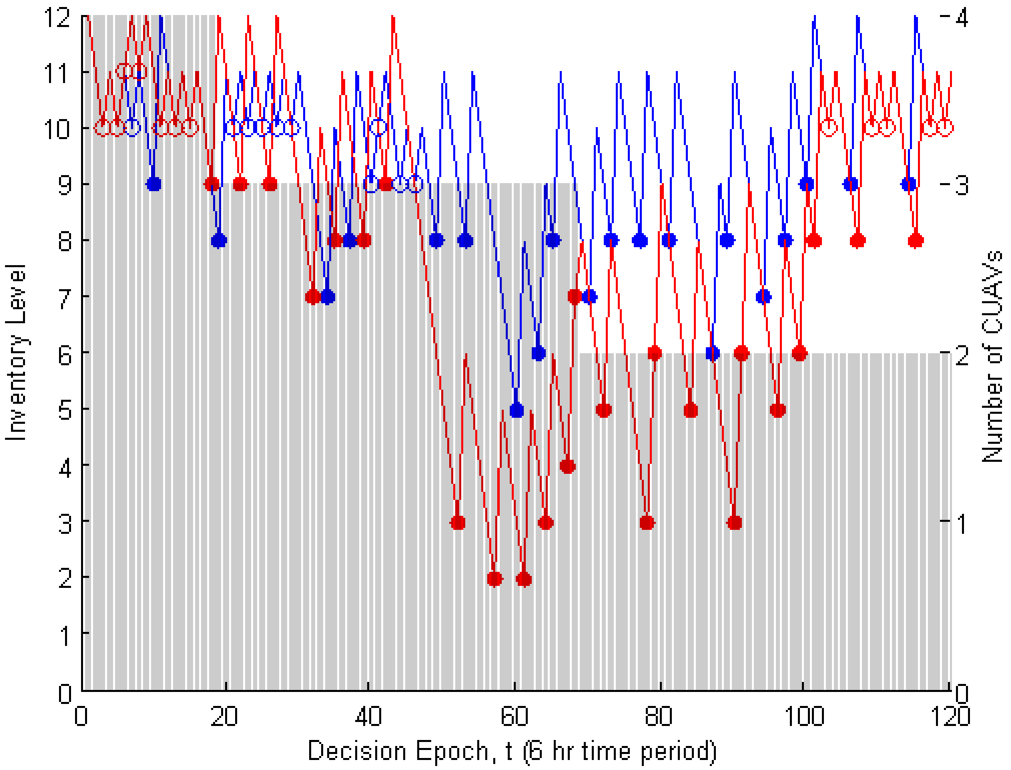


Figure 11. 2-COP 30-day sample path of an IVAPI policy (with second order basis functions)

Figure 12 shows the sample path when the myopic policy is used in the simulation. In this particular simulation, COP 1's inventory level follows the exact same path as COP 2's, and is hidden beneath the red lines and circles. All four CUAVs are destroyed in the first 15 decision epochs, a product of the fact that under the myopic policy CUAVs are deployed even when inventory levels and threat conditions (indicated by Map 2) are quite high.

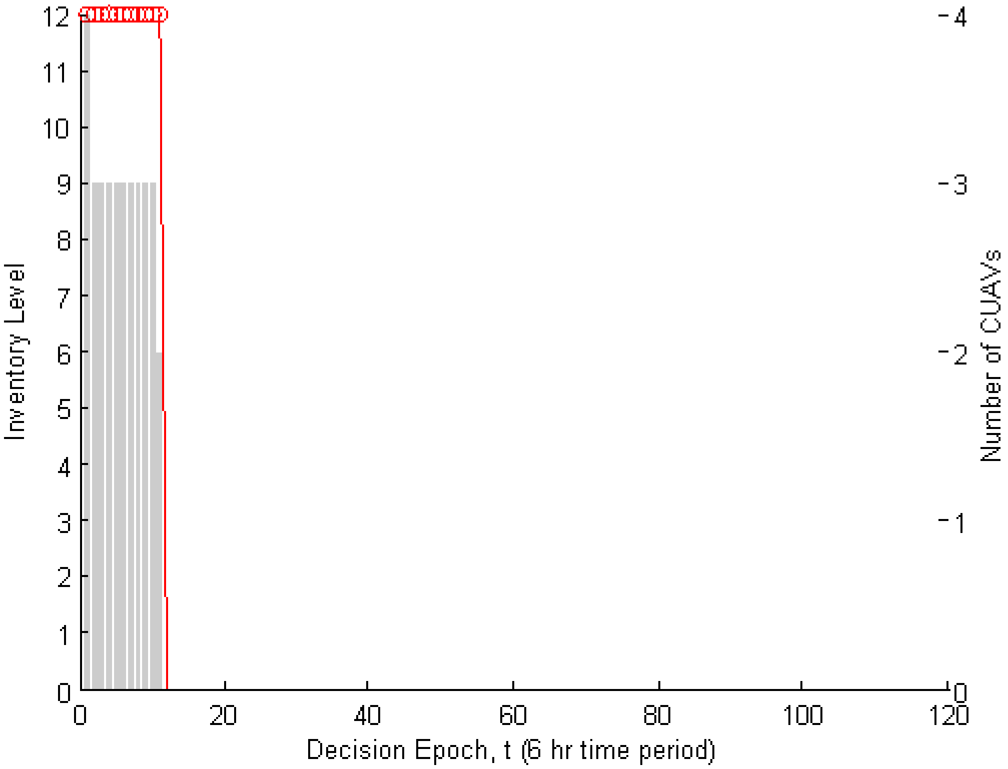


Figure 12. 2-COP 30-day sample path of the myopic policy

The results in this chapter suggests that IVAPI (when used with a second order model with indicator variables and interaction terms) is promising. With the 2-COP computational example fully explored, we turn to the 3-COP problem instance. The results from the 3-COP problem instance are used to evaluate the consistency of the ADP and basis functions for future use in the 12-COP problem instance.

Appendix C. Computational Example: 3-COP

Optimal Policy

Table 25 displays the optimal policy under Map 1 when one CUAV is available and inventory at COP 3 is one. We observe that for a majority of the inventory levels at COPs 1 and 2, the optimal decision is to send a CUAV to COP 3. This is understandable as COP 3's inventory is quite low. However, as inventory at COP 1 drops to one or inventory at COP 2 drops to one, the single CUAV is sent to the depleted COP. When more than one COP is depleted, the CUAV is deployed to the COP with the higher ψ-value. For example, ψ_{21} is greater than ψ_{31}. Therefore when inventory at COPs 2 and 3 are one and inventory at COP 1 is higher than three, we observe that the CUAV is deployed to COP 2 rather than COP 3. When all COPs are at inventory levels of one, the CUAV is sent to COP 1, the COP with the highest ψ-value under Map 1.

Table 25. 3-COP Optimal Policy, Map = 1, Vehicles = 1, Inventory at COP 3 = 1

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)
	2	(0,1,0)	(0,1,0)	(1,0,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	3	(0,0,1)	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	4	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	5	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	6	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	7	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	8	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	9	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	10	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	11	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	12	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)

Table 26 displays the optimal policy under Map 1 when three CUAVs are available and inventory at COP 3 is six. When inventory at COP 1 is low and inventory at COP 2 is high, the optimal action is to send three CUAVs to COP 1. However, as COP 1's inventory increases, the three available CUAVs are split between COP 1 and COP 2. Not until inventory at COP 1 is 12 does the optimal policy change to sending all three CUAVs to COP 2. When inventory levels at both COPs 1 and 2 are high, the CUAVs are evenly split between the COPs or divided amongst the COPs. In every state, all three available CUAVs are being deployed.

Table 26. 3-COP Optimal Policy, Map = 1, Vehicles = 3, Inventory at COP 3 = 6

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,2,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	2	(1,2,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	3	(1,2,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	4	(1,2,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	5	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	6	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	7	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	8	(1,2,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,0,1)
	9	(0,3,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,0,1)
	10	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,1,1)	(1,1,1)	(1,0,2)
	11	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,1,1)	(1,1,1)	(1,0,2)
	12	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,2,1)	(0,2,1)	(0,1,2)	(0,1,2)	(0,0,3)

Table 27 displays the optimal policy under Map 1 when six CUAVs are available and inventory at COP 3 is 12. With a crew of three, only three CUAVs can be deployed in a single time epoch. When inventory levels at both COP 1 and COP 2 are high, only two CUAVs are deployed. This reflects the high inventory level at COP 3 and the desire to reserve CUAVs for later use when the COPs have lower inventory levels. Otherwise, when inventory at either COP is low and the inventory level at the other COP is high, all three CUAVs are sent to resupply the more depleted COP. When inventory levels for COPs 1 and 2 are about equal, the three CUAVs are split between the two COPs.

Table 27. 3-COP Optimal Policy, Map = 1, Vehicles = 6, Inventory at COP 3 = 12

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	2	(1,2,0)	(2,1,0)	(2,1,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	3	(1,2,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	4	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	5	(0,3,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	6	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	7	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	8	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(3,0,0)
	9	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,0,0)
	10	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(2,1,0)	(1,1,0)	(2,0,0)
	11	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,1,0)	(1,1,0)	(1,0,0)
	12	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(0,2,0)	(1,1,0)	(0,1,0)	(1,0,0)

Table 28 displays the optimal policy under Map 2 when one CUAV is available and inventory at COP 3 is one. The optimal policy here is (with three exceptions) to send a CUAV to COP 3 to avoid the GLOC penalty. While under Map 1, the optimal decision when supply at COP 1 was low was to resupply COP 1. However, due to the fact that COP 3 has the highest ψ-value in Map 2, the optimal action usually is to send a CUAV to COP 3.

Table 28. 3-COP Optimal Policy, Map = 2, Vehicles = 1, Inventory at COP 3 = 1

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	2	(0,0,1)	(1,0,0)	(1,0,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	3	(0,0,1)	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	4	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	5	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	6	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	7	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	8	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	9	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	10	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	11	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	12	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)

Table 29 displays the optimal policy under Map 2 when three CUAVs are available and inventory at COP 3 is six. Unlike under Map 1, there are many instances where the optimal policy is to deploy no CUAVs due to the lower ψ-values in Map 2. When inventory at COP 2 and COP 3 are high, the optimal policy to send CUAVs to COP 3 (again, a reflection of the high COP 3 ψ-value under Map 2). When inventory at COP 1 or COP 2 becomes one, the depleted COPs are replenished instead of serving COP 3.

Table 29. 3-COP Optimal Policy, Map = 2, Vehicles = 3, Inventory at COP 3 = 6

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,2,0)	(2,1,0)	(2,1,0)	(2,0,1)	(2,0,1)	(2,0,1)	(2,0,1)	(2,0,1)	(2,0,1)	(2,0,1)	(2,0,1)	(2,0,1)
	2	(0,3,0)	(0,2,1)	(0,2,1)	(0,1,2)	(0,0,3)	(0,0,3)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	3	(0,3,0)	(0,2,1)	(0,1,2)	(0,0,3)	(0,0,3)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	4	(0,3,0)	(0,2,1)	(0,1,2)	(0,0,3)	(0,0,3)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	5	(0,3,0)	(0,1,2)	(0,0,3)	(0,0,3)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	6	(0,3,0)	(0,1,2)	(0,0,3)	(0,0,3)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	7	(0,3,0)	(0,1,2)	(0,0,3)	(0,0,3)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	8	(0,3,0)	(0,1,2)	(0,0,3)	(0,0,3)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,1)
	9	(0,3,0)	(0,1,2)	(0,0,3)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	10	(0,3,0)	(0,1,2)	(0,0,3)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,3)	(0,0,3)	(0,0,3)
	11	(0,3,0)	(0,0,1)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)
	12	(0,3,0)	(0,0,1)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)

Finally, Table 30 displays the optimal policy under Map 2 when six CUAVs are available, and inventory at COP 3 is maximized at 12. Unless inventory at COP 1 is one or inventory at COP 2 is one or two, the optimal policy is to deploy no CUAVs. When COP 1 or 2 has an inventory of one, all three CUAVs are deployed to the depleted COP(s) to avoid a GLOC penalty.

Table 30. 3-COP Optimal Policy, Map = 2, Vehicles = 6, Inventory at COP 3 = 12

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	2	(0,3,0)	(0,2,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	3	(0,3,0)	(0,1,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	4	(0,3,0)	(0,1,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	5	(0,3,0)	(0,1,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	6	(0,3,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	7	(0,3,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	8	(0,3,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	9	(0,3,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	10	(0,3,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	11	(0,3,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)
	12	(0,3,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)

The optimal policy tables reflect three ideas. First, as the number of available CUAVs increases, more risk is taken and additional CUAVs are deployed. Second, Map 1 is preferred over Map 2 due to larger ψ-values in Map 1 than in Map 2 for the first two COPs. Finally, as inventory level at a COP increases, a CUAV is less likely to be deployed to that COP. Again, these optimal policy tables support the same general conclusions from the 2-COP example.

Myopic Policy

Table 31 displays the myopic policy under Map 1 when one vehicle is available and COP 3's inventory is one. Unless COP 1 or COP 2 has an inventory level of one, the optimal action is to send a CUAV to COP 3 to avoid the GLOC penalty. When inventory levels at COP 1 or COP 2 are one, the single available CUAV is sent to the depleted COP. When both COPs have inventory levels of one, the CUAV is deployed to COP 1, a reflection of the higher ψ-value for COP 1 under Map 1.

Table 31. 3-COP Myopic Policy, Map = 1, Vehicles = 1, Inventory at COP 3 = 1

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)
	2	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	3	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	4	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	5	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	6	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	7	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	8	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	9	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	10	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	11	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	12	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)

Table 32 displays the myopic policy under Map 1 when three vehicles are available and COP 3's inventory is six. When inventory at COP 1 is less than seven, usually the optimal policy is to send a CUAV to COP 1. However, as COP 1's inventory increases, the three CUAVs are split between COP 1 and COP 2. When inventory at COP 1 is maximized, the CUAVs are sent to COP 2 and/or COP 3.

Table 32. 3-COP Myopic Policy, Map = 1, Vehicles = 3, Inventory at COP 3 = 6

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,2,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	2	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	3	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	4	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	5	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	6	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	7	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	8	(0,3,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,0,1)
	9	(0,3,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,0,1)
	10	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,1,1)	(1,1,1)	(1,0,2)
	11	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,1,1)	(1,1,1)	(1,0,2)
	12	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,2,1)	(0,2,1)	(0,1,2)	(0,1,2)	(0,0,3)

Table 33 displays the myopic policy under Map 1 when six vehicles are available and COP 3's inventory is maximized at 12. The myopic policy for this state is very similar to the previous policy (map = 1, CUAVs = 3, and inventory at COP 3 = 6). This is a reflection of the fact that the number of CUAVs which can be deployed in a single time epoch is limited by the number of crews. This difference between the policies is found when inventory at both COP 1 and COP 2 are higher than nine. In Table 33, a CUAV is deployed to COP 3 less than in Table 32. This myopic policy tends to spread CUAVs out over more COPs at high levels of inventory rather than sending all three CUAVs to a single COP.

Table 33. 3-COP Myopic Policy, Map = 1, Vehicles = 6, Inventory at COP 3 = 12

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(1,2,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	2	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	3	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	4	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	5	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	6	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	7	(0,3,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	8	(0,3,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(3,0,0)
	9	(0,3,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)	(2,1,0)
	10	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(2,1,0)
	11	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,1,1)	(1,1,1)
	12	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(1,1,1)	(1,1,1)

Table 34 displays the myopic policy under Map 2 when one vehicle is available and COP 3's inventory is one. The optimal action for all states is to send the CUAV to COP 3. This is a product of the high ψ-value in Map 2 for COP 3 and the low inventory level at COP 3.

Table 34. 3-COP Myopic Policy, Map = 2, Vehicles = 1, Inventory at COP 3 = 1

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	2	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	3	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	4	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	5	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	6	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	7	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	8	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	9	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	10	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	11	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)
	12	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)	(0,0,1)

Table 35 displays the myopic policy under Map 2 when three vehicles are available and COP 3's inventory is six. With the exception of inventory levels of one at COP 1 or COP 2, the optimal policy is to deploy the CUAV to COP 3. When inventory at COP 1 or COP 2 is one, CUAVs are deployed to the depleted COP.

Table 35. 3-COP Myopic Policy, Map = 2, Vehicles = 3, Inventory at COP 3 = 6

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	2	(0,3,0)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)
	3	(0,3,0)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)
	4	(0,3,0)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)
	5	(0,3,0)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)
	6	(0,3,0)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)
	7	(0,3,0)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)
	8	(0,3,0)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)
	9	(0,3,0)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)
	10	(0,3,0)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)
	11	(0,3,0)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)
	12	(0,3,0)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)	(0,0,3)

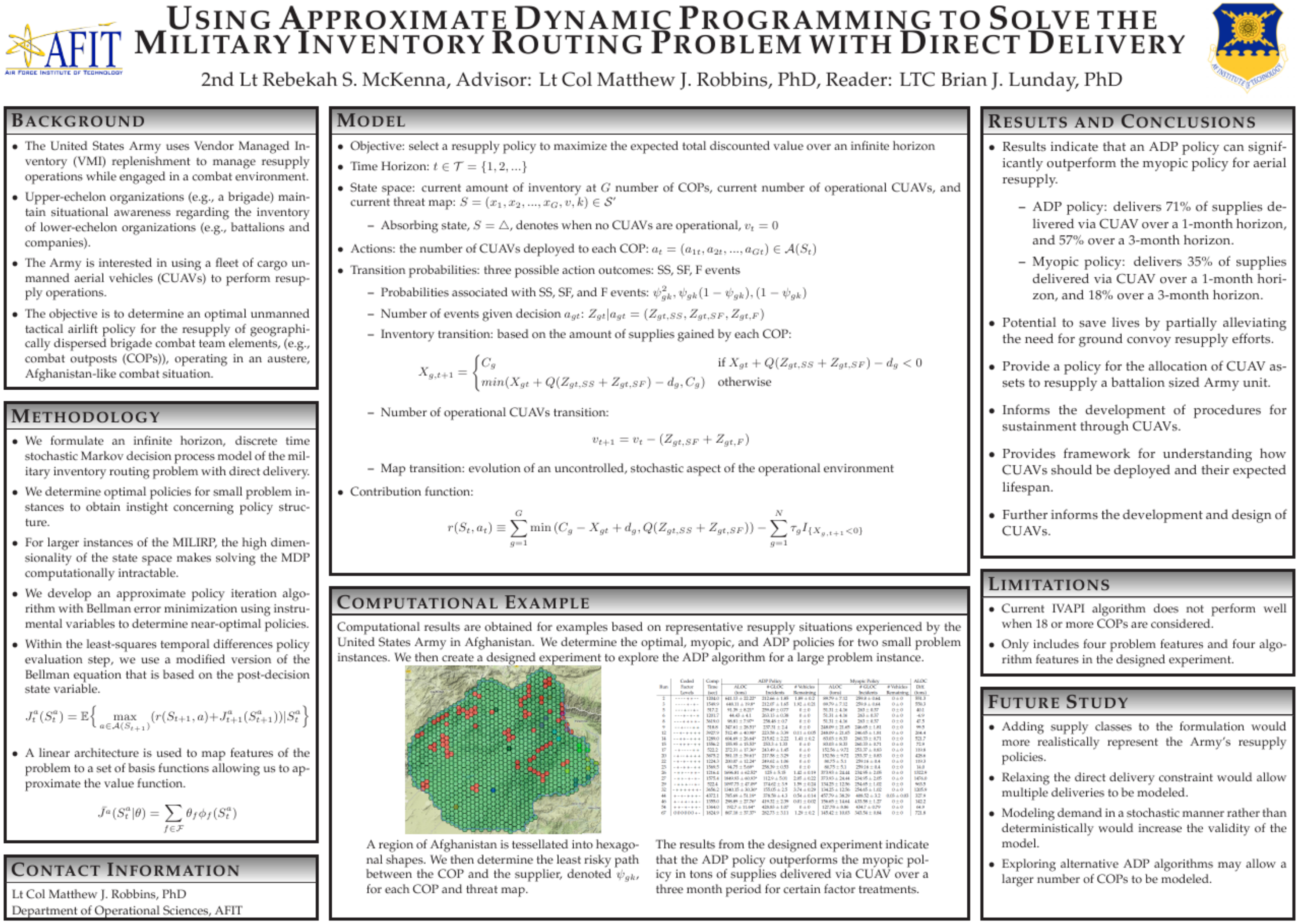
Finally, Table 36 displays the myopic policy under Map 2 when six vehicles are available and COP 3's inventory is maximized at 12. When inventory at COP 2 is below eight, the optimal policy is usually to send three CUAVs to COP 2. When inventory at COP 2 increases, the CUAVs are split between the COPs or all sent to COP 1 (when inventory at COP 1 is low).

Table 36. 3-COP Myopic Policy, Map = 2, Vehicles = 6, Inventory at COP 3 = 12

		Inventory COP 2											
		1	2	3	4	5	6	7	8	9	10	11	12
Inventory COP 1	1	(2,1,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)	(3,0,0)
	2	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)
	3	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)
	4	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)
	5	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)
	6	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)
	7	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)
	8	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(3,0,0)
	9	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,1,0)	(2,0,1)
	10	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(2,1,0)	(2,0,1)
	11	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,1,1)	(1,1,1)
	12	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,3,0)	(0,2,1)	(0,2,1)	(1,1,1)	(1,1,1)

In every case for the myopic policy, all available CUAVs (limited by the number of crews) are deployed in a manner that maximizes the reward for the single time epoch. The myopic policy conclusions for the 3-COP problem parallels the conclusions from the 2-COP myopic policy.

Appendix D. Storyboard



Using Approximate Dynamic Programming to Solve the Military Inventory Routing Problem with Direct Delivery

2nd Lt Rebekah S. McKenna, Advisor: Lt Col Matthew J. Robbins, PhD, Reader: LTC Brian J. Lunday, PhD

Background

- The United States Army uses Vendor Managed Inventory (VMI) replenishment to manage resupply operations while engaged in a combat environment.
- Upper-echelon organizations (e.g., a brigade) maintain situational awareness regarding the inventory of lower-echelon organizations (e.g., battalions and companies).
- The Army is interested in using a fleet of cargo unmanned aerial vehicles (CUAVs) to perform resupply operations.
- The objective is to determine an optimal unmanned tactical airlift policy for the resupply of geographically dispersed brigade combat team elements, (e.g., combat outposts (COPs)), operating in an austere, Afghanistan-like combat situation.

Methodology

- We formulate an infinite horizon, discrete time stochastic Markov decision process model of the military inventory routing problem with direct delivery.
- We determine optimal policies for small problem instances to obtain insight concerning policy structure.
- For larger instances of the MILIRP, the high dimensionality of the state space makes solving the MDP computationally intractable.
- We develop an approximate policy iteration algorithm with Bellman error minimization using instrumental variables to determine near-optimal policies.
- Within the least-squares temporal differences policy evaluation step, we use a modified version of the Bellman equation that is based on the post-decision state variable.

$$J^a_t(S^a_t) = \mathbb{E}\left\{ \max_{a \in \mathcal{A}(S_{t+1})} (r(S_{t+1}, a) + J^a_{t+1}(S^a_{t+1})) | S^a_t \right\}$$

- A linear architecture is used to map features of the problem to a set of basis functions allowing us to approximate the value function.

$$\bar{J}^a(S^a_t|\theta) = \sum_{f \in \mathcal{F}} \theta_f \phi_f(S^a_t)$$

Contact Information

Lt Col Matthew J. Robbins, PhD
Department of Operational Sciences, AFIT

Model

- Objective: select a resupply policy to maximize the expected total discounted value over an infinite horizon
- Time Horizon: $t \in \mathcal{T} = \{1, 2, ...\}$
- State space: current amount of inventory at G number of COPs, current number of operational CUAVs, and current threat map: $S = (x_1, x_2, ..., x_G, v, k) \in \mathcal{S}'$
  - Absorbing state, $S = \triangle$, denotes when no CUAVs are operational, $v_t = 0$
- Actions: the number of CUAVs deployed to each COP: $a_t = (a_{1t}, a_{2t}, ..., a_{Gt}) \in \mathcal{A}(S_t)$
- Transition probabilities: three possible action outcomes: SS, SF, F events
  - Probabilities associated with SS, SF, and F events: $\psi^2_{gk}, \psi_{gk}(1 - \psi_{gk}), (1 - \psi_{gk})$
  - Number of events given decision a_{gt}: $Z_{gt}|a_{gt} = (Z_{gt,SS}, Z_{gt,SF}, Z_{gt,F})$
  - Inventory transition: based on the amount of supplies gained by each COP:

$$X_{g,t+1} = \begin{cases} C_g & \text{if } X_{gt} + Q(Z_{gt,SS} + Z_{gt,SF}) - d_g < 0 \\ min(X_{gt} + Q(Z_{gt,SS} + Z_{gt,SF}) - d_g, C_g) & \text{otherwise} \end{cases}$$

  - Number of operational CUAVs transition:

$$v_{t+1} = v_t - (Z_{gt,SF} + Z_{gt,F})$$

  - Map transition: evolution of an uncontrolled, stochastic aspect of the operational environment
- Contribution function:

$$r(S_t, a_t) \equiv \sum_{g=1}^{G} \min\left(C_g - X_{gt} + d_g, Q(Z_{gt,SS} + Z_{gt,SF})\right) - \sum_{g=1}^{N} \tau_g I_{\{X_{g,t+1} < 0\}}$$

Computational Example

Computational results are obtained for examples based on representative resupply situations experienced by the United States Army in Afghanistan. We determine the optimal, myopic, and ADP policies for two small problem instances. We then create a designed experiment to explore the ADP algorithm for a large problem instance.

A region of Afghanistan is tessellated into hexagonal shapes. We then determine the least risky path between the COP and the supplier, denoted ψ_{gk}, for each COP and threat map.

Run	Coded Factor Levels	Comp Time (sec)	ADP Policy			Myopic Policy			ALOC Diff. (tons)
			ALOC (tons)	# GLOC Incidents	# Vehicles Remaining	ALOC (tons)	# GLOC Incidents	# Vehicles Remaining	
1	[illegible]	1204.0	641.13 ± 22.22*	212.66 ± 1.88	1.89 ± 0.2	89.79 ± 7.12	259.8 ± 0.64	0 ± 0	551.3
3	[illegible]	1548.9	648.11 ± 19.8*	212.07 ± 1.65	1.92 ± 0.21	89.79 ± 7.12	259.8 ± 0.64	0 ± 0	558.3
5	[illegible]	517.2	91.39 ± 8.21*	259.49 ± 0.77	0 ± 0	51.31 ± 4.36	263 ± 0.37	0 ± 0	40.1
6	[illegible]	1201.7	44.43 ± 4.1	263.13 ± 0.38	0 ± 0	51.31 ± 4.36	263 ± 0.37	0 ± 0	-6.9
8	[illegible]	3619.0	98.81 ± 7.97*	258.48 ± 0.7	0 ± 0	51.31 ± 4.36	263 ± 0.37	0 ± 0	47.5
9	[illegible]	518.8	347.61 ± 28.51*	237.31 ± 2.4	0 ± 0	248.09 ± 21.65	246.65 ± 1.81	0 ± 0	99.5
12	[illegible]	3927.9	512.49 ± 40.98*	223.56 ± 3.39	0.11 ± 0.05	248.09 ± 21.65	246.65 ± 1.81	0 ± 0	264.4
14	[illegible]	1289.0	604.89 ± 26.84*	215.82 ± 2.22	1.41 ± 0.2	83.03 ± 8.33	260.33 ± 0.71	0 ± 0	521.7
15	[illegible]	1186.2	155.93 ± 15.53*	253.3 ± 1.33	0 ± 0	83.03 ± 8.33	260.33 ± 0.71	0 ± 0	72.9
17	[illegible]	522.2	272.31 ± 17.36*	243.49 ± 1.45	0 ± 0	152.56 ± 9.72	253.37 ± 0.83	0 ± 0	119.8
20	[illegible]	3873.2	581.35 ± 39.61*	217.58 ± 3.29	0 ± 0	152.56 ± 9.72	253.37 ± 0.83	0 ± 0	428.8
22	[illegible]	1224.3	208.07 ± 12.24*	249.62 ± 1.06	0 ± 0	88.75 ± 5.1	259.16 ± 0.4	0 ± 0	119.3
23	[illegible]	1568.5	94.73 ± 5.69*	258.39 ± 0.53	0 ± 0	88.75 ± 5.1	259.16 ± 0.4	0 ± 0	6.0
26	[illegible]	1216.4	1696.81 ± 62.82*	125 ± 5.18	1.42 ± 0.19	373.93 ± 24.44	234.95 ± 2.05	0 ± 0	1322.9
27	[illegible]	1575.4	1849.93 ± 60.93*	112.9 ± 5.01	2.05 ± 0.22	373.93 ± 24.44	234.95 ± 2.05	0 ± 0	1476.0
29	[illegible]	522.4	1097.73 ± 47.09*	174.62 ± 3.9	1.59 ± 0.24	134.25 ± 12.56	254.65 ± 1.02	0 ± 0	963.5
32	[illegible]	3656.2	1340.15 ± 30.36*	155.05 ± 2.5	3.74 ± 0.29	134.25 ± 12.56	254.65 ± 1.02	0 ± 0	1205.9
44	[illegible]	4372.1	785.69 ± 51.19*	378.59 ± 4.3	0.54 ± 0.14	457.79 ± 38.39	408.52 ± 3.2	0.05 ± 0.03	327.9
46	[illegible]	1555.0	298.89 ± 27.76*	419.32 ± 2.39	0.01 ± 0.02	156.65 ± 14.64	433.58 ± 1.27	0 ± 0	142.2
54	[illegible]	1364.0	192.7 ± 11.64*	428.83 ± 1.07	0 ± 0	127.79 ± 8.86	434.7 ± 0.79	0 ± 0	64.9
67	[illegible]	1624.9	867.18 ± 37.37*	282.73 ± 3.11	1.29 ± 0.2	145.42 ± 10.83	343.54 ± 8.84	0 ± 0	721.8

The results from the designed experiment indicate that the ADP policy outperforms the myopic policy in tons of supplies delivered via CUAV over a three month period for certain factor treatments.

Results and Conclusions

- Results indicate that an ADP policy can significantly outperform the myopic policy for aerial resupply.
  - ADP policy: delivers 71% of supplies delivered via CUAV over a 1-month horizon, and 57% over a 3-month horizon.
  - Myopic policy: delivers 35% of supplies delivered via CUAV over a 1-month horizon, and 18% over a 3-month horizon.
- Potential to save lives by partially alleviating the need for ground convoy resupply efforts.
- Provide a policy for the allocation of CUAV assets to resupply a battalion sized Army unit.
- Informs the development of procedures for sustainment through CUAVs.
- Provides framework for understanding how CUAVs should be deployed and their expected lifespan.
- Further informs the development and design of CUAVs.

Limitations

- Current IVAPI algorithm does not perform well when 18 or more COPs are considered.
- Only includes four problem features and four algorithm features in the designed experiment.

Future Study

- Adding supply classes to the formulation would more realistically represent the Army's resupply policies.
- Relaxing the direct delivery constraint would allow multiple deliveries to be modeled.
- Modeling demand in a stochastic manner rather than deterministically would increase the validity of the model.
- Exploring alternative ADP algorithms may allow a larger number of COPs to be modeled.

Bibliography

1. Barnes-Schuster, Dawn, & Bassok, Yehuda. 1997. Direct shipping and the dynamic single-depot/multi-retailer inventory system. *Eurpoean Journal of Operational Research*, 509–518.

2. Barr, Richard S., Golden, Bruce L., Kelly, James P., Resende, Mauricio G.C., & Stewart, William R., Jr. 1995. Designing and reporting on computational experi–ments with heuristic methods. *Journal of Heuristics*, **1**(1), 9–32.

3. Bertazzi, Luca. 2008. Analysis of Direct Shipping Policies in an Inventory-Routing Problem with Discrete Shipping Times. *Management Science*, **54**(4), 748 – 762.

4. Bertazzi, Luca, Bosco, Adamo, Guerriero, Francesca, & Lagana, Demetrio. 2013. A stochastic inventory routing problem with stock-out. *Transportation Research: Part C*, **27**, 89–107.

5. Bradtke, Steven J., Barto, Andrew G., & Kaelbling, Pack. 1996. Linear least–-squares algorithms for temporal difference learning. 22–33.

6. Coelho, Leandro C., & Laporte, Gilbert. 2013. The exact solution of several classes of inventory-routing problems. *Computers & Operations Research*, **40**(2), 558 – 565.

7. Coelho, Leandro C., Cordeau, Jean-Franois, & Laporte, Gilbert. 2012. Thirty Years of Inventory Routing. *Transportation Science*, **48**(1), 1–19.

8. Department of the Army. 2010. Army Field Manual: Brigade Combat Team No. 3-90.6. *Army Knowledge Online.*

9. Department of the Army. 2012. Cargo Unmanned Aircraft System (UAS) Concept of Operations.

10. General Dynamics Information Technology. 2010. *Future Modular Force Resupply Mission for Unmanned Aircraft Systems (UAS).* General Dynamics Information Technology.

11. Godfrey, Gregory A., & Powell, Warren B. 2001. An Adaptive, Distribution-Free Algorithm for the Newsvendor Problem with Censored Demands, with Applications to Inventory and Distribution. *Management Science*, **47**(8), 1101.

12. Kleywegt, Anton J., Nori, Vijay S., & Savelsbergh, Martin W. P. 2002. The Stochastic Inventory Routing Problem with Direct Deliveries. *Transportation Science*, **36**(1), 94.

13. Kleywegt, Anton J., Nori, Vijay S., & Savelsbergh, Martin W. P. 2004. Dy–namic Programming Approximations for a Stochastic Inventory Routing Problem. *Transportation Science*, **38**(1), 42 – 70.

14. Lagoudakis, Michail G, & Parr, Ronald. 2003. Least-squares policy iteration. *The Journal of Machine Learning Research*, **4**, 1107–1149.

15. Lamothe, Dan. *Robotic helicopter completes Afghanistan mission, back in U.S.* `http://www.washingtonpost.com/news/checkpoint/wp/2014/07/25/robotic-helicopter-completes-afghanistan-mission-back-in-u-s/`. Accessed: 2015-02-18.

16. Lockheed Martin. *K-MAX Unmanned Arcraft System.* `http://www.lockheedmartin.com/content/dam/lockheed/data/ms2/documents/K-MAX-brochure.pdf`. Accessed: 2014-10-18.

17. Lockheed Martin. *Unmanned K-MAX Operations in Afghanistan.* `https://www.youtube.com/watch?v=s-mr5I657GU`. Accessed: 2015-02-19.

18. McCormack, Ian. 2014. *The Military Inventory Routing Problem with Direct Delivery.* M.Phil. thesis, Air Force Institute of Technology.

19. Mu, S., Fu, Z., Lysgaard, J., & Eglese, R. 2010. Disruption Management of the Vehicle Routing Problem with Vehicle Breakdown. *Journal of the Operational Research Society*, **62**, 742–749.

20. Novoa, Clara, & Storer, Robert. 2009. An Approximate Dynamic Programming Approach for the Vehicle Routing Problem with Stochastic Demands. *European Journal of Operational Research*, **196**, 509–515.

21. Powell, Warren B. 2011. *Approximate Dynamic Programming: Solving the Curses of Dimensionality.* 2 edn. John Wiley & Sons, Inc.

22. Puterman, Martin L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* John Wiley & Sons, Inc.

23. Scott, Warren R., Powell, Warren B., & Moazehi, Somayeh. 2014. Least Squares Policy Iteration with Instrumental Variables vs. Direct Policy Search: Comparison Against Optimal Benchmarks Using Energy Storage.

24. Stevenson, Beth. *Unmanned K-Max to Undergo Further Testing.* `http://www.flightglobal.com/news/articles/`. Accessed: 2015-02-18.

25. Topaloglu, Huseyin, & Powell, Warren B. 2003. An algorithm for approximating piecewise linear concave functions from sample gradients. *Operations Research Letters*, **31**(1), 66.

26. Tsitsiklis, John N., & Sutton, Richard. 1994. Asynchronous Stochastic Approximation and Q-Learning. 185–202.

27. United States Department of Defense. 2009. FY 2009-2034 Unmanned Systems Integrated Roadmap. April.

28. Williams, Jason. 2010. *Unmanned Tactical Airlift: A Business Case Study.* M.Phil. thesis, Air Force Institute of Technology.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704–0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704–0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD–MM–YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From — To)*
26–03–2015	Master's Thesis	SEP 2013 — MAR 2015

4. TITLE AND SUBTITLE Using Approximate Dynamic Programming to Solve the Military Inventory Routing Problem with Direct Delivery	**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) McKenna, Rebekah S., Second Lieutenant, USAF	**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/ENS) 2950 Hobson Way WPAFB OH 45433-7765	**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT-ENS-MS-15-M-140
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) TRADOC Capability Manager for Unmanned Aircraft Systems Deputy, TCM-UAS Mr. Glenn A. Rizzi 453 Novosel Street Fort Rucker, AL 36362 glenn.a.rizzi.civ@mail.mil	**10. SPONSOR/MONITOR'S ACRONYM(S)** TCM-UAS **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution Statement A.
Approved for Public Release; distribution unlimited.

13. SUPPLEMENTARY NOTES

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

The United States Army uses Vendor Managed Inventory (VMI) replenishment to manage resupply operations while engaged in a combat environment; upper-echelon organizations (e.g., a brigade) maintain situational awareness regarding the inventory of lower-echelon organizations (e.g., battalions and companies). The Army is interested in using a fleet of cargo unmanned aerial vehicles (CUAVs) to perform resupply operations. We formulate an infinite horizon, discrete time stochastic Markov decision process model of the military inventory routing problem with direct delivery, the objective of which is to determine an optimal unmanned tactical airlift policy for the resupply of geographically dispersed brigade combat team elements operating in an austere, Afghanistan-like combat situation. An approximate policy iteration algorithm with Bellman error minimization using instrumental variables is applied to determine near-optimal policies. Within the least-squares temporal differences policy evaluation step, we use a modified version of the Bellman equation that is based on the post-decision state variable. Computational results are obtained for examples based on representative resupply situations experienced by the United States Army in Afghanistan.

15. SUBJECT TERMS

inventory routing problem, Markov decision process, approximate dynamic programming

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. JD Robbins, AFIT/ENS
U	U	U	U	122	**19b. TELEPHONE NUMBER** *(include area code)* (937) 255-3636, x4539 matthew.robbins@afit.edu